

TRANSMITTAL LETTER TO THE UNITED STATES  
DESIGNATED/ELECTED OFFICE (DO/EO/US)  
CONCERNING A FILING UNDER 35 U.S.C. 371

U.S. APPLICATION NO. (If known, see 37 CFR 1.5)

09/623068

INTERNATIONAL APPLICATION NO.  
PCT/US99/04376INTERNATIONAL FILING DATE  
26 February 1999PRIORITY DATE CLAIMED  
26 February 1998

TITLE OF INVENTION

TWO-DIMENSIONAL LINKAGE STUDY TECHNIQUES

532 Rec'd PCT/PTC 26 AUG 2000

APPLICANT(S) FOR DO/EO/US

MCGINNIS, Ralph, Evan

MCGINNIS, Robert Owen

Applicant herewith submits to the United States Designated/Elected Office (DO/EO/US) the following items and other information:

1. ☒ This is a **FIRST** submission of items concerning a filing under 35 U.S.C. 371.
2. ☐ This is a **SECOND** or **SUBSEQUENT** submission of items concerning a filing under 35 U.S.C. 371.
3. ☐ This express request to begin national examination procedures (35 U.S.C. 371(f)) at any time rather than delay examination until the expiration of the applicable time limit set in 35 U.S.C. 371(b) and PCT Articles 22 and 39(1).
4. ☒ A proper Demand for International Preliminary Examination was made by the 19th month from the earliest claimed priority date.
5. ☒ A copy of the International Application as filed (35 U.S.C. 371(c)(2))
  - a. ☐ is transmitted herewith (required only if not transmitted by the International Bureau).
  - b. ☐ has been transmitted by the International Bureau.
  - c. ☒ is not required, as the application was filed in the United States Receiving Office (RO/US).
6. ☐ A translation of the International Application into English (35 U.S.C. 371(c)(2)).
7. ☒ Amendments to the claims of the International Application under PCT Article 19 (35 U.S.C. 371(c)(3))
  - a. ☐ are transmitted herewith (required only if not transmitted by the International Bureau).
  - b. ☒ have been transmitted by the International Bureau.
  - c. ☐ have not been made; however, the time limit for making such amendments has NOT expired.
  - d. ☐ have not been made and will not be made.
8. ☐ A translation of the amendments to the claims under PCT Article 19 (35 U.S.C. 371(c)(3)).
9. ☒ An oath or declaration of the inventor(s) (35 U.S.C. 371(c)(4)).
10. ☐ A translation of the annexes to the International Preliminary Examination Report under PCT Article 36 (35 U.S.C. 371(c)(5)).

**Items 11. to 16. below concern document(s) or information included:**

11. ☐ An Information Disclosure Statement under 37 CFR 1.97 and 1.98.
12. ☐ An assignment document for recording. A separate cover sheet in compliance with 37 CFR 3.28 and 3.31 is included.
13. ☐ A **FIRST** preliminary amendment.  
☐ A **SECOND** or **SUBSEQUENT** preliminary amendment.
14. ☐ A substitute specification.
15. ☐ A change of power of attorney and/or address letter.
16. ☒ Other items or information:

Two signed small entity statements. A signed letter (2 pages long) requesting amendments. 7 Replacement pages, each replacement page with header "PCT/US99/04376(U.S.National Stage Entry Aug. 2000)". Replacement pages are as follows: background replacement page number 5, background replacement page number 6, background replacement page number 7, background replacement page number 8, specification replacement page number 38, specification replacement page number 43, specification replacement page number 46. Express Mail Mailing Certificate, Return Receipt Postcard, Cenwest check for \$48 (small entity Basic National Fee). Statement under Article 19(1).

09/623068

PCT/US99/04376

09/623068 09/623068

17. ☒ The following fees are submitted:

BASIC NATIONAL FEE (37 CFR 1.492 (a) (1) - (5)) :

Neither international preliminary examination fee (37 CFR 1.482)  
nor international search fee (37 CFR 1.445(a)(2)) paid to USPTO  
and International Search Report not prepared by the EPO or JPO . . . . . \$970.00

International preliminary examination fee (37 CFR 1.482) not paid to  
USPTO but International Search Report prepared by the EPO or JPO . . . . . \$840.00

International preliminary examination fee (37 CFR 1.482) not paid to USPTO but  
international search fee (37 CFR 1.445(a)(2)) paid to USPTO . . . . . \$690.00

International preliminary examination fee paid to USPTO (37 CFR 1.482)  
but all claims did not satisfy provisions of PCT Article 33(1)-(4) . . . . . \$670.00

International preliminary examination fee paid to USPTO (37 CFR 1.482)  
and all claims satisfied provisions of PCT Article 33(1)-(4) . . . . . \$96.00

ENTER APPROPRIATE BASIC FEE AMOUNT =

\$ 96.00

Surcharge of \$130.00 for furnishing the oath or declaration later than ☐ 20 ☐ 30  
months from the earliest claimed priority date (37 CFR 1.492(e)).

\$

CLAIMS	NUMBER FILED	NUMBER EXTRA	RATE
Total claims	20-20 =	0	X \$18.00
Independent claims	3-3 =	0	X \$78.00

\$

\$

MULTIPLE DEPENDENT CLAIM(S) (if applicable) + \$260.00

\$

TOTAL OF ABOVE CALCULATIONS =

\$ 96.00

Reduction of 1/2 for filing by small entity, if applicable. A Small Entity Statement  
must also be filed (Note 37 CFR 1.9, 1.27, 1.28).

\$

48.00

SUBTOTAL =

\$ 48.00

Processing fee of \$130.00 for furnishing the English translation later than ☐ 20 ☐ 30  
months from the earliest claimed priority date (37 CFR 1.492(f)).

\$

TOTAL NATIONAL FEE =

\$ 48.00

Fee for recording the enclosed assignment (37 CFR 1.21(h)). The assignment must be  
accompanied by an appropriate cover sheet (37 CFR 3.28, 3.31). \$40.00 per property

\$

TOTAL FEES ENCLOSED =

\$ 48.00

Amount to be

refunded:

\$

charged:

\$ 48.00

a. ☒ A check in the amount of \$ 48.00 to cover the above fees is enclosed.

b. ☐ Please charge my Deposit Account No. \_\_\_\_\_ in the amount of \$ \_\_\_\_\_ to cover the above fees.  
A duplicate copy of this sheet is enclosed.

c. ☐ The Commissioner is hereby authorized to charge any additional fees which may be required, or credit any  
overpayment to Deposit Account No. \_\_\_\_\_. A duplicate copy of this sheet is enclosed.

**NOTE:** Where an appropriate time limit under 37 CFR 1.494 or 1.495 has not been met, a petition to revive (37 CFR 1.137(a) or (b)) must be filed and granted to restore the application to pending status.

SEND ALL CORRESPONDENCE TO:

Robert Owen McGinnis  
1575 West Kagy Blvd.  
Bozeman, Montana 59715

Robert McGinnis

SIGNATURE:

Robert O. McGinnis

NAME

44, 232

REGISTRATION NUMBER

## Two Dimensional Linkage Study Techniques

### Technical Field

Versions of the present invention are in the field of molecular biology, some versions are specifically in the area of finding the chromosomal location of genes that cause genetic characteristics such as human disease.

### Background

#### Introduction

Conventional linkage study techniques have limited power to localize trait causing genes ( trait causing polymorphisms ) of modest effect, such as many human disease polymorphisms. The two-dimensional linkage study techniques of this application are powerful new techniques for localizing genes (polymorphisms) especially of modest effect.

#### Chromosomes, heredity, genes, markers and alleles

Chromosomes are large molecules that carry the information for the inheritance of physical (genetic) characteristics or traits. In human beings for example, parents pass a copy of half of their chromosomes to their offspring during reproduction. By doing this, each parent passes some of his or her physical characteristics to his or her offspring. Any chromosome of a living creature is made of a large string-like molecule of DNA. Chromosomes are essentially very long strings of DNA. Genes are small pieces of a chromosome that cause or determine inherited genetic characteristics. (In this application, the term gene means a polymorphism that determines a genetic characteristic; the term does not mean an entire gene structure with a promoter region, introns, etc..) Markers are any segment of DNA on a chromosome which can be identified and whose chromosomal location is known (at least to some extent). Markers are like milestones along the very long string-like molecule of DNA which makes up a chromosome. Both a gene and a marker can come in different forms on different chromosomes. These different forms are known as different alleles and when a gene or marker comes in different forms it is said to be "polymorphic". For example, a bi-allelic marker comes in two (bi) different forms.

#### Linkage

If a gene allele and a marker allele occur as part of the genetic makeup of individuals more frequently than would be expected on the basis of chance, then it is possible to infer that the gene and the marker are linked. If a gene allele and a marker allele are inherited together more frequently than would be expected if the gene and the marker were on different chromosomes, then it is possible to infer that the gene and the marker are linked. Linkage of a gene and a marker usually occurs because the gene and the marker are close together on a chromosome. There are different degrees of linkage. Establishing linkage, especially strong linkage, between a gene and a marker can be very valuable. This is especially true if the precise location and other characteristics of the gene are not known. By establishing linkage, especially strong linkage, between a known marker and an unknown gene it is possible to locate the gene near to the chromosomal location of the known marker. This can be very

valuable if the gene is an important gene, such as a disease causing gene, and can help cure the disease.

### Linkage Studies

Linkage studies are a method of establishing linkage between a marker and a gene or genes. Linkage studies are used to statistically correlate the occurrence of a genetic characteristic such as a disease (caused by a gene or genes) with a marker on a chromosome. One way this is done is by statistically correlating a specific allele of a marker with a genetic characteristic for a set of individuals by showing that individuals with the characteristic inherit the marker allele more often than individuals without the characteristic. The set of individuals is usually referred to as a sample of individuals. An example of a sample of individuals are people with a disease and similar people (matched controls) without the disease. Another example of a sample of individuals is a group of people, some of whom have the same disease; each of the people in the group being related to one or more of the other people in the group (i.e. families, sibships, pedigrees). The presence or absence of a marker allele in the chromosomal DNA of each individual is usually determined by genotype data at the marker for each individual.

There are different types of linkage study techniques, using different types of samples and different statistical measures of the correlation of a marker and a genetic characteristic. One example of a type of linkage study technique is the affected sib pair (ASP) test. Another example is the transmission disequilibrium test (TDT), which is an association based linkage test. This is a dynamic, changing area within the field of human genetics.

### Linkage Studies and the "Scanning" of Chromosomal Regions

There are significant advantages in using several markers simultaneously to perform a linkage study with a genetic characteristic and a sample of individuals, especially when the relative positions of the markers on a chromosome are known. Such a linkage study allows searching for statistical evidence of linkage between markers in one or more regions of a chromosome or chromosomes and the gene or genes that determine the genetic characteristic. The results of the study for each marker can then be compared with the results for other markers, knowing the relative chromosomal positions of all the markers in the study. In this way, regions of a chromosome or even whole chromosomes can be "scanned" for evidence of linkage to a gene or genes causing a genetic characteristic. The relative positions of markers on chromosomes of a species of creatures is given by various kinds of chromosomal maps for the species. (There are several different kinds of marker maps, i.e. physical maps, genetic maps, radiation hybrid maps, etc.)

### Sets of Markers for Linkage Studies and "Scanning" Chromosomes

An appropriate set of markers from a region of a chromosome can be chosen so that the region can be "scanned" for evidence of linkage of markers in the region to a gene or genes that cause a genetic characteristic. As explained above, this scanning is done by using the markers in linkage studies. Strong positive evidence for linkage of the markers (from the scanned chromosomal region) to a gene or genes responsible for a characteristic or trait is strong evidence that a trait-causing gene or genes is located within the chromosomal region.

Conventional Techniques for Choosing Sets of Markers to Scan Chromosomes with Linkage Studies

Conventional techniques choose sets of markers to scan a chromosomal region by choosing markers according to each marker's chromosomal location within the region. In a set of microsatellite markers described in 1994 for use in linkage studies, the markers were approximately evenly spaced, with average spacing between markers being 13 centiMorgans. The markers were distributed approximately evenly across the entire human genome (all human chromosomes) and were also selected because genotype data at the markers for individuals could be obtained by a semi-automated method.<sup>1</sup> A recent (1998) linkage study of the disease schizophrenia used a set of 310 microsatellite markers distributed approximately evenly across the entire human genome with average spacing of 11 centiMorgans between markers.<sup>2</sup> In a recent (1998) simulation of linkage studies to defend the practice of two-stage genome scanning, markers were spaced evenly every 10 cM (centimorgans) in an initial, sparser, first stage scan and evenly every 1 cM in a followup, denser, second stage scan.<sup>3</sup> Following up positive linkage study results from chromosomal regions in a sparse, first stage scan with a second, denser scan that focuses on studying the regions with positive first-stage results is a common technique. In these conventional studies, as is common, markers were chosen to be about evenly spaced across the chromosomal regions studied. In this manner, as is conventional, a one dimensional structure such as an entire genome, a chromosome or a region of a chromosome is "covered" by markers in order to scan the entire genome, chromosome or chromosomal region with a linkage study. (These conventional techniques<sup>1, 2, 3</sup> are not admitted to be prior art by their mention in this background.) (There is a possibly confusing, double meaning, of the term "marker map" It should be noted that a set of markers distributed along a chromosomal region, chromosome, or genome for linkage studies is also sometimes referred to as a "marker map" for use in chromosomal scanning by linkage studies. In addition, chromosomal or genetic maps of markers are also referred to as "marker maps".)

Conventional Techniques for Choosing Sets of Markers to Scan Chromosomal Regions are Essentially One Dimensional

Because DNA is a stringlike molecule, a chromosomal region(s), chromosome(s) and genome are essentially one dimensional in terms of the chromosomal location of markers and genes. As has been stated, conventional linkage study techniques scan a chromosomal region(s), chromosome(s) or genome by using markers distributed approximately evenly along the length of the chromosomal region(s), chromosome(s) or genome respectively. These conventional techniques focus primarily on the chromosomal location of markers used in a scan. These conventional techniques have an essentially one dimensional perspective.

Population Frequency of Marker Alleles and Gene Alleles

As described, chromosomal location of each marker is an important and unique characteristic of each marker and marker allele. Another characteristic of each polymorphic marker and each of the marker's

<sup>1</sup> Reed, et.al.: Chromosome-specific microsatellite sets for fluorescence-based, semi-automated genome mapping. Nature Genetics, July 1994; vol. 7: pp. 390-395.

<sup>2</sup> Levinson, et.al.: Genome Scan of Schizophrenia. Am J Psychiatry, June 1998; vol. 155: pp. 741-750.

<sup>3</sup> Kruglyak, et. al.: Linkage Thresholds for Two-stage Genome Scans. Am J Hum Genet, 1998, vol. 62: pp. 994-996.

alleles is the population frequency of each marker allele. A population is a group (usually a large group) of individuals. A population frequency of a particular marker allele is the proportion of individual chromosomes in a population in which the particular marker occurs as the particular marker allele. For any bi-allelic marker, knowing the least common allele frequency of the marker establishes both of the allele frequencies of the marker. This is because the two allele frequencies of a bi-allelic marker sum to 1. Each gene allele also has a population allele frequency or allele frequency for short. Thus, each gene allele has a particular chromosomal location and allele frequency (for a particular population). In the case of an unknown gene, the gene's chromosomal location and allele frequencies are not specifically known.

#### Marker Allele Population Frequency in Conventional Linkage Study Scans

It is important to note that little attention was paid to the population allele frequencies of the markers used in the conventional linkage scans cited above. In the two studies cited above under conventional scanning techniques<sup>1,2</sup>, marker allele frequency is referred to only peripherally as average marker heterozygosity, which is related to average marker allele frequency and the number of alleles (2, 3, 4, 5, etc.) at each marker. In the simulated scan cited above<sup>3</sup>, *the markers are stipulated to have four alleles that all have exactly the same allele frequency of 0.25 (heterozygosity 0.75). It is important to note that while the chromosomal location of the markers in all these conventional scans was systematically varied over the entire genome (all the human chromosomes), nothing was said about systematically varying the allele frequencies of the markers in any of the scans.* This is typical of conventional linkage study scans of genomes, chromosomes and chromosomal regions.

#### A Conventional View Of Bi-allelic Markers And Linkage Studies

We cite here a well known reference that discusses the conventional view of bi-allelic marker usefulness in linkage scans of chromosomes. In 1997 Kruglyak carried out computer simulations of the "information content" of markers that are part of various different marker maps.<sup>4</sup> For bi-allelic markers his results showed that the optimum allele frequencies for bi-allelic markers used in linkage studies is 0.5/0.5 in order to achieve the greatest information content. However, allele frequency patterns other than the optimum 0.5/0.5 for bi-allelic markers gave acceptable levels of information content depending on the density of the marker map (or set of markers) chosen for the linkage study.

There are some important observations regarding this reference.<sup>4</sup> First, *there is no advantage noted in this reference for choosing bi-allelic markers so that the set of chosen markers (or marker map) used for linkage studies is such that the markers systematically vary in allele frequency.* Thus, just as in the recent conventional linkage study scans cited above, there is no definite thought to using markers of systematically varying allele frequencies. The greatest information content is given by bi-allelic markers with allele frequencies close to the optimum of 0.5/0.5. Given the density of reasonably polymorphic SNPs predicted in this reference, at least one every 1 kb or 1,000 per cM, it is probable that even for quite dense maps, there will be so many acceptable SNPs available, that all of the SNPs in an appropriate marker map could have the optimum allele frequencies of approximately

<sup>4</sup> Kruglyak: The use of a genetic map of biallelic markers in linkage studies. Nature Genetics, September 1997, vol.17, pp. 21-24.

0.5/0.5. Secondly, bi-allelic markers with lower least common allele frequencies, less than 0.3(0.7/0.3) or 0.2(0.8/0.2), are viewed unfavorably for linkage studies in this reference. Thirdly, the early version of the criterion of "information content" of markers used in this reference was based on sib pair analysis and the later, current version of the criterion, does not depend on any particular test for linkage.<sup>5, 6</sup> Thus, the criterion of information content in this reference, has never specifically employed the TDT (transmission disequilibrium test) or any association based test, whereas the two-dimensional linkage study techniques of this application are based on a completely different perspective of using association based tests. (This reference<sup>1</sup> is not admitted to be prior art with respect to the present invention by it's mention in this background.)

Increased Power of the TDT (transmission disequilibrium test)

Characteristics of a new type of linkage test, the TDT (transmission disequilibrium test), were described in 1993. The inventor, R.E.McGinnis, was one of the authors of this reference.<sup>7</sup> In 1996, Risch and Merikangas argued that conventional linkage analysis has limited power to detect genes of modest effect. And Risch and Merikangas attempted to illustrate the increased power of association based linkage tests such as the TDT over other types of conventional linkage tests.<sup>8</sup> However, Risch and Merikangas' analysis was criticized by Muller-Myhsok and Abel as being based on the optimal assumption that the analyzed allele was the disease allele itself. Muller-Myhsok and Abel concluded that researchers should be aware that the power of association studies such as the TDT can be greatly diminished in more common, less optimal situations.<sup>9</sup> In their response to Muller-Myshok and Abels' letter, Risch and Merikangas essentially agreed with the logic of Muller-Myshok and Abels' criticism. Risch and Merikangas stated that to a large extent, the expectation with respect to linkage disequilibrium across the genome is uncharted territory.<sup>10</sup> (None of the references in this paragraph<sup>7,8, 9,10</sup> is admitted to being prior art with respect to the present invention by their mention in this background.)

More Detailed Studies of the Power of the TDT

The inventor, R.E.McGinnis, has done extensive investigations on the power of the TDT. His observations and calculations of the increased power of the TDT in many situations have been

<sup>5</sup> Kruglyak, et. al.: Complete Multipoint Sib-Pair Analysis of Qualitative and Quantitative Traits. Am J Hum Genet. 1995, vol. 57: pp. 439-454.

<sup>6</sup> Kruglyak, et. al.: Parametric and Nonparametric Linkage Analysis: A Unified Multipoint Approach. Am J Hum Genet, 1996, vol. 58, pp. 1347- 1363.

<sup>7</sup> Spielman, R.S., McGinnis, R.E., Ewens, W.J.: Transmission Test for Linkage Disequilibrium: The Insulin Gene Region and Insulin-dependent Diabetes Mellitus(IDDM). Am J Hum Genet, 1993, vol. 52, pp. 506-516.

<sup>8</sup> Risch, N. and Merikangas, K.: The Future of Genetic Studies of Complex Human Diseases. Science. 13 September 1996, vol. 273, pp. 1516-1517.

<sup>9</sup> Muller-Myshok, B. and Abel, L.: Technical Comments: The Future of Complex Diseases. Science. 28 February 1997, vol. 275, pp. 1328-1329.

<sup>10</sup> Risch, N. and Merikangas, K.: Technical Comments: The Future of Complex Diseases. Science. 28 February 1997, vol. 275, p. 1330.

published.<sup>11</sup> In this paper a general framework for determining the power of the TDT in many different situations is presented. The analysis of Risch and Merikangas<sup>8</sup> and others is shown by the inventor to be a special case of his general framework. His observations and calculations published in this paper have shown that the TDT has increased power in more common, less optimal situations as well as the less common, optimal situation cited by Muller-Myshok and Abel<sup>9</sup>. As opposed to the observation of Muller-Myshok and Abel, the inventor's calculations indicate that association tests such as the TDT have increased power in typical situations even when the ratio m/p departs significantly from unity and, or the linkage disequilibrium between the analyzed (marker) allele and disease polymorphism is only half its maximum possible value. The inventor arrived at these conclusions independently and did not derive them from others.

**A Major Conclusion Drawn by the Inventor about the TDT and Linkage Studies: Using Bi-allelic Markers of Systematically Varying Allele Frequencies Increases the Power of Linkage Studies Using the TDT**

The inventor's calculations and observations about the increased power of the TDT in more common, less optimal situations led him to the conclusion that the power of linkage studies using the TDT is greatly increased under some conditions. Under some conditions, the power of the TDT in a linkage study using bi-allelic markers is greatly increased when each of one or more of the bi-allelic markers used in the study fulfill two criteria: (1) the allele frequencies of each of the one or more of the bi-allelic markers are similar (but not necessarily the same, or even approximately the same) as the allele frequencies of an unknown bi-allelic gene causing a disease under study; and (2) each of the one or more bi-allelic markers is in some degree of linkage disequilibrium with the gene. Thus for a typical linkage study using bi-allelic markers and the TDT, ***to increase the likelihood of conditions occurring that increase the power of the TDT in the linkage study, the bi-allelic markers used in the study are chosen so that the least common allele frequencies of the markers vary systematically over a range or subrange of least common allele frequency.*** This major conclusion of the inventor's research is quoted directly from his unpublished manuscript that was included with previously filed U.S. Provisional Patent Applications: "This example is typical and highlights perhaps the most important finding of this paper; namely the importance of using bi-allelic markers with heterozygosity similar to that of a bi-allelic disease gene. Indeed, since a majority of susceptibility loci may be bi-allelic, the judicious use of bi-allelic markers of both high, medium and low heterozygosity may be crucial in order to detect and replicate linkages to loci conferring modest disease risk." (page 25) (In this context the phrase "bi-allelic markers with heterozygosity similar to that of a bi-allelic disease gene" is essentially equivalent to "bi-allelic markers with individual allele frequencies similar to those of a bi-allelic disease gene" and "bi-allelic markers of both high, medium and low heterozygosity" is essentially equivalent to the phrase "bi-allelic markers whose least common individual allele frequencies are high, medium and low".)

<sup>11</sup> McGinnis, R.E.: Hidden Linkage: Comparison of the affected sib pair (ASP) test and transmission disequilibrium test (TDT). *Annals of Human Genetics*, 1998, vol. 62, pp. 159-179.



1 Systematically Varying Both Marker Chromosomal Location and Marker Allele Frequency of Markers in  
2 Linkage Studies

3 The inventor's calculations and observations have demonstrated the increased power of the TDT in  
4 more common, less optimal situations when a bi-allelic marker and bi-allelic gene have (1) similar but  
5 not identical allele frequencies and (2) the marker and gene are in some degree of linkage  
6 disequilibrium. Thus, for a typical linkage study using bi-allelic markers and the TDT, ***to increase the***  
7 ***likelihood of both criteria (1) and (2) occurring for one or more markers, so as to increase the***  
8 ***power of the TDT in the linkage study, the bi-allelic markers used in the study are chosen so***  
9 ***that the least common allele frequencies of the markers vary systematically over a range or***  
10 ***subrange of least common allele frequency AND the chromosomal location of the markers vary***  
11 ***systematically over one or more chromosomes or chromosomal regions. And the bi-allelic***  
12 ***markers are chosen so that the markers' chromosomal locations and least common allele***  
13 ***frequencies vary systematically in an essentially independent manner.***

14 Two-dimensional Linkage Study Techniques

15 As has been stated, conventional linkage study scanning techniques use markers that are distributed  
16 approximately evenly in the dimension of chromosomal location. These conventional, one dimensional,  
17 scanning techniques focus primarily on the chromosomal location of markers used in a scan and give  
18 little attention to the dimension of allele frequency.<sup>1, 2, 3</sup>

19 One of the main implications of the inventor's work is to use a set of bi-allelic markers for a typical  
20 linkage study using the TDT (or other association-based linkage test) wherein the chromosomal  
21 locations and least common allele frequencies of the markers in the set systematically vary in an  
22 essentially independent manner over the dimensions of chromosomal location and least common allele  
23 frequency respectively. This is equivalent to using a set of bi-allelic markers for a linkage study scan  
24 wherein the set of markers systematically scan or "cover" a two-dimensional region having dimensions  
25 of chromosomal location and least common allele frequency. (Such a two-dimensional region can be  
26 thought of as an area in an x-y plot or a group of squares on a chessboard.)

27 In addition, the inventor's calculations and observations indicate that bi-allelic markers having least  
28 common allele frequencies less than 0.3, 0.2 or even less than 0.1 have an important place in linkage  
29 studies using association based linkage tests. This is markedly different than Kruglyak's information  
30 content evaluation of bi-allelic markers for use in linkage studies, in which bi-allelic markers with least  
31 common allele frequencies less than 0.3 or 0.2 are viewed unfavorably.<sup>4</sup>

32 *In addition, the two-dimensional linkage study techniques do not necessarily favor using markers in a*  
33 *scan that are about evenly spaced along a chromosome as in the conventional techniques. This is*  
34 ***because conventional techniques suffer from a kind of one dimensional view or lack of depth***  
35 ***perception. In the conventional techniques, a marker can look very close to a gene's location in***  
36 ***terms of chromosomal location, but the marker can be very far from the gene's location in the***  
37 ***new two dimensional view used by versions of the invention.***

38 ***It is as if the conventional 1D techniques look at a chessboard from on edge. Markers and a***  
39 ***gene which are on different squares of the board, but in the same column of squares, look very***

close to each other when the board is looked at from on edge. But when the board is looked at from the top in 2D, two dimensions, markers which looked very close to each other and the gene before (when looking from on edge) can be seen to be very far from the gene.

Further Implications of the Two-dimensional Linkage Study Perspective

These two-dimensional techniques work when multiple genes cause a genetic characteristic and are effective in searching for these genes. A two-dimensional bi-allelic marker "covering" or scanning approach also increases the power of linkage studies using other association based linkage tests such as the AFBACmethod, the haplotype relative risk (HRR) method<sup>12</sup>, and comparison of marker allele frequencies in disease cases and unrelated controls<sup>13</sup>. These references<sup>12, 13</sup> are not admitted to being prior art with respect to the present invention by their mention in this background.)

Patents That May Be Helpful In Starting A Search Of The Background

Some patents that are in the same general areas as versions of the invention are cited here: US Patent Number 5,667,976 Solid supports for nucleic acid hybridization assays. Published International Application WO 98/20165 Biallelic Markers. Published International Application WO 98/07887 Methods for treating bipolar mood disorder associated with markers on chromosome 18 p. US Patent Number 5,552,270 Methods of DNA sequencing by hybridization based on optimizing concentration of matrix-bound oligonucleotide and device for carrying out same. No patent in this paragraph is admitted to being prior art with respect to the present invention by it's mention in this background.

<sup>12</sup> Falk CT and Rubenstein P: Haplotype relative risks: an easy reliable way to construct a proper control sample for risk calculations. Annals of Human Genetics, 1987, vol. 51, pp. 227-233.

<sup>13</sup> Bell GI, Horita S and Karam JH: A polymorphic locus near the human insulin gene is associated with insulin-dependent diabetes mellitus. Diabetes, 1984, vol 33, pp. 176-183.

## Two-Dimensional Linkage Study Techniques

### Brief Description of Some Concepts Used By Versions of the Invention

Versions of the present invention make use of the novel concept of systematically covering a region on a two-dimensional map similar to an x-y graph with bi-allelic markers. The x axis on this map is the chromosomal location dimension and the y axis of the map is the least common allele frequency dimension. This two-dimensional map is called a CL-F map in this application. (CL stands for chromosomal location and F stands for least common allele frequency.) Each point on a CL-F map has two coordinates: a chromosomal location coordinate and a frequency coordinate. A point on a CL-F map is called a CL-F point.

Any one bi-allelic polymorphism (marker or gene) is viewed as being located at a particular CL-F point on a CL-F map. The chromosomal location of the polymorphism is the chromosomal location coordinate of the point. And the least common allele frequency of the polymorphism is the frequency coordinate of the point. The chromosomal location coordinate of a CL-F point is given in units of centiMorgans or base pairs or an equivalent thereof and the least common allele frequency coordinate of a CL-F point is given in units between 0 and 0.5 inclusive, such as 0.2.

Distances between any two CL-F points on a CL-F map are given in terms of two numbers: chromosomal location distance and frequency distance. The first number is the distance in the horizontal, chromosomal location direction. This first number is the chromosomal location distance. The second number is the distance in the vertical, frequency direction. This second number is the frequency distance. For example, the CL-F distance  $\delta$  is given by two numbers  $\delta_{CL}$  (chromosomal location distance) and  $\delta_F$  (frequency distance). This is represented as  $\delta = [\delta_{CL} \delta_F]$ .

The "clustering" of bi-allelic markers near a particular CL-F point is discussed in terms of the number of markers within a particular CL-F distance of the point. For example, if each of N bi-allelic markers is separated from the point by a CL-F distance of less than or equal to  $\delta$ , then the point is said to be N covered by the markers to within the distance  $\delta$ . (N being an integer number.)

A region on a CL-F map is called a CL-F region. A CL-F region is a collection of one or more CL-F points. Some systematic methods of covering a CL-F region with bi-allelic markers are discussed in terms of the number of markers that are near each point in the region. For example, if each CL-F point in a CL-F region is N covered to within a CL-F distance  $\delta$  by a subset of a set (or group) of bi-allelic markers, then the region is said to be N covered by the set (or group) of bi-allelic markers to within the distance  $\delta$ .

A set (or group) of bi-allelic markers that cover a CL-F region or a CL-F point is referred to as a set (or group) of bi-allelic covering markers in this application.

The inventor discovered that when a bi-allelic marker and a bi-allelic gene are located close together on a CL-F map, then the power of association based linkage tests to detect linkage disequilibrium between the marker and a trait-causing gene (when present) increases greatly. Systematically covering a CL-F region that is the location of an unknown trait-causing bi-allelic gene with bi-allelic covering markers, therefore greatly increases the power of association based linkage tests to detect linkage disequilibrium (when present) between one or more of the covering markers and the gene.

PCT/US 99/04376

IPEA/US 24 MAY 2000

10

1 A CL-F matrix is a matrix of rectangular cells of the same length and the same width on a CL-F map.  
2 Stipulations that a certain number of covering markers are placed in each cell of the matrix is a method  
3 of illustrating particular types of systematic covering of a CL-F region with covering markers.  
4 The evidence for linkage obtained from two-dimensional linkage studies is essentially two-dimensional  
5 in nature and it is possible to use this two-dimensional information by essentially graphing quantitative  
6 evidence for linkage as a function of position in the x-y plane. For example, if quantitative evidence for  
7 linkage is represented in the z dimension of a typical three-dimensional x-y-z plot, wherein the x and y  
8 dimensions are chromosomal location and least common allele frequency respectively, then it is  
9 possible to conceptualize evidence for linkage as occurring in a "hump" or "humps" in the z dimension.  
10 And it is possible to analyze the data to find the CL-F location (in the x-y plane) of the peak(s) of this  
11 "hump(s)", thus helping to localize a trait causing gene to the CL-F locale of the peak(s) of the  
12 "hump(s)".  
13 Versions of the invention also make use of multi-allelic genes and/or markers. It is always possible to  
14 combine the alleles of a multi-allelic polymorphism (marker or gene) so that the polymorphism acts  
15 mathematically like it is a bi-allelic polymorphism. In effect, it is always possible to mathematically  
16 transform a multi-allelic marker or gene to act bi-allelic. Similarly, two or more markers can always be  
17 mathematically combined to form a mathematical marker that acts like a single bi-allelic marker. And  
18 two or more genes can always be mathematically combined to form a mathematical gene that acts like  
19 a single bi-allelic gene. In this application a mathematical bi-allelic marker formed mathematically from  
20 one or more markers is called a bi-allelic marker equivalent or BME; and a mathematical bi-allelic gene  
21 formed mathematically from one or more genes is called a bi-allelic gene equivalent or BGE.  
22 The term true marker or gene is used to distinguish a marker or gene in the ordinary sense from a bi-  
23 allelic marker equivalent (BME) or bi-allelic gene equivalent (BGE). The term true allele is used to  
24 distinguish an allele in the ordinary sense from a mathematical allele of a BME or BGE. A mathematical  
25 allele of a BME or BGE is referred to as an allele equivalent. An allele equivalent is a combination of  
26 one or more true alleles or one or more haplotypes.  
27 Versions of the invention make use of genes and/or markers, which are not exactly bi-allelic. These  
28 genes or markers are approximately bi-allelic. A gene or marker that is approximately bi-allelic almost  
29 always occurs in one of two allele forms, however, very rarely it occurs in a different allele form.  
30 Various versions of the invention are for genotyping individuals at markers which systematically cover  
31 CL-F regions or for obtaining sample allele frequency data (such as from pooled DNA) for a sample of  
32 individuals for markers which systematically cover CL-F regions. Various versions of the invention are  
33 for oligonucleotides used for genotyping individuals at markers which systematically cover CL-F regions  
34 or are for obtaining sample allele frequency data (such as from pooled DNA) for a sample of individuals  
35 for markers which systematically cover CL-F regions.

#### Definitions

37  
38  
39 For the purposes of the description and claims the terms used herein will have their generally accepted  
40 definition unless otherwise specified.

AMENDED SHEET

SCANNED # 14

1 The term **creature** means any organism that is living or was alive at one time. This includes both plants  
2 and animals.

3 The term **species** is used in it's broadest sense and includes but is not limited to : 1)biological(genetic)  
4 species,2) paleospecies (successional species), 3) taxonomic (morphological ; phenetic) species  
5 including species hybrids such as mules, 4) microspecies ( agamospecies) 5) biosystematic species(  
6 coenospecies,ecosystem species)

7 A **genetic characteristic** is an observable or inferable inherited genetic characteristic or inherited  
8 genetic trait including a biochemical or biophysical genetic trait, for example an inherited disease is a  
9 genetic characteristic, a predisposition to an inherited disease is a genetic characteristic. A phenotypic  
10 characteristic, phenotypic property or character is a genetic characteristic.

11 In this application, **the term gene** means a polymorphism that takes on one or more allele forms and  
12 which causes or determines an inherited genetic characteristic or genetic trait. **The term gene** does not  
13 mean an entire gene structure with a promoter region, a terminator region, introns, and other parts of an  
14 entire gene structure. In this application the term gene means a polymorphism that determines or  
15 causes an inherited genetic characteristic and that is part of an entire gene structure in some cases.  
16 Each **genetic characteristic** of a creature is **determined** by one or more of the creature's **genes**,  
17 wherein the term gene is defined as above.

18 A **segment** is a segment of a chromosome.

19 A **subrange** is a subrange of the least common allele frequency range 0 to 0.5 inclusive.

20 The **width** of a subrange is the difference between the upper and lower limits of the subrange. For  
21 example, the width of the subrange 0.1 to 0.4 is  $0.4 - 0.1 = 0.3$ .

22 A **chromosomal location-least common allele frequency map** is a two-dimensional plot (similar to  
23 an x-y graph) wherein the vertical axis(y axis) represents least common allele frequency and the  
24 horizontal axis(x axis) represents chromosomal location. A chromosomal location-least common allele  
25 frequency map is referred to as a **CL-F map**.

26 **Points on a CL-F map are referred to as CL-F points.** Points on a CL-F map have a chromosomal  
27 location coordinate and a least common allele frequency coordinate. CL-F points represent possible  
28 chromosomal location and least common allele frequency values for individual bi-allelic markers and  
29 genes. Any particular point on a CL-F map is directly opposite a value on the map's least common  
30 allele frequency axis(y axis) and is directly opposite a value on the map's chromosomal location axis(x  
31 axis). These two values are the two coordinates of the particular point: (1) the chromosomal location  
32 coordinate and (2) the least common allele frequency coordinate. A marker or gene located at a  
33 particular point on a CL-F map is physically located at the chromosomal location given by the  
34 chromosomal location coordinate of the point and the marker or gene's least common allele frequency  
35 is the least common allele frequency coordinate of the point. These two coordinates are designated by  
36 the term ( x, y ) wherein x is the value of the chromosomal location coordinate and y is the value of the  
37 least common allele frequency coordinate.

38 A **particular CL-F map may be large or small.** For example it is possible for the chromosomal  
39 location coordinates of CL-F points on a particular CL-F map to range over an entire chromosome ( for  
40 example human chromosome number 6). Alternatively it is possible for the chromosomal location



1 **A CL-F region** is a group of CL-F points. A CL-F region is a region that is or can be represented on a  
 2 CL-F map. A particular CL-F region may be large or small. For example the chromosomal location  
 3 coordinates of CL-F points in a particular CL-F region can range over an entire chromosome ( for  
 4 example human chromosome number 6). Alternatively the chromosomal location coordinates of CL-F  
 5 points in a particular CL-F region can range over more than one chromosome, for example all the  
 6 human chromosomes, human chromosomes numbers 1 through 22 and X and Y. Similarly the  
 7 chromosomal location coordinates of CL-F points in a particular CL-F region can range over all the  
 8 chromosomes of a species under study. Alternatively, the chromosomal location coordinates of CL-F  
 9 points in a particular CL-F region can range over only a small segment of chromosome, for example a  
 10 segment of length 100,000 bp or less. Similarly the least common allele frequency coordinates of CL-F  
 11 points in a particular CL-F region can range over the entire least common allele frequency range 0 to  
 12 0.5. Alternatively the least common allele frequency coordinates of CL-F points in a particular CL-F  
 13 region can range over only a very small subrange, for example the subrange 0.1 to 0.2 or less.

14 **The length of a CL-F region** is the largest chromosomal location distance between any two CL-F  
 15 points in the region.

16 **The width of a CL-F region** is the largest frequency distance between any two CL-F points in the  
 17 region.

18 **A CL-F region that is path connected** is contiguous and it is possible to draw a continuous path  
 19 between any two points, wherein each point in the path is also in the region.

20 **If a CL-F region is said to be systematically covered by two or more bi-allelic covering markers**  
 21 then each point in the region is within a small CL-F distance of one or more of the covering markers,  
 22 wherein the magnitude of the small CL-F distance is such that there is increased power of an  
 23 association based linkage test to detect evidence for linkage between one or more covering markers  
 24 and a gene that is located at a point in the CL-F region, when linkage disequilibrium is present between  
 25 the gene and one or more of the covering markers.

26 **If a CL-F region is said to be N covered to within a CL-F distance  $\delta$  by one or more covering**  
 27 **markers** then each point in the region is N covered to within the CL-F distance  $\delta$  by the one or more  
 28 covering markers. Wherein N is an integer greater than or equal to one.

29 **If a CL-F region is said to be N covered to within a CL-F distance of about (or approximately)  $\delta$**   
 30 **by one or more covering markers** then each point in the region is N covered to within the CL-F  
 31 distance of about (or approximately)  $\delta$  by the one or more covering markers. Wherein N is an integer  
 32 greater than or equal to one.

33 **The CL-F distance  $\delta$  is known as the covering distance** if a CL-F point or CL-F region is N covered  
 34 to within a CL-F distance  $\delta$ .

35 **A CL-F covering distance  $\delta$  has two components:** (1) a chromosomal location distance usually  
 36 denoted by  $\delta_{CL}$  and (2) a least common allele frequency distance (abbreviated as frequency distance)  
 37 usually denoted by  $\delta_F$ , i.e.  $\delta = [\delta_{CL}, \delta_F]$ .

38 **The length of a group of covering markers** is determined as follows. The absolute chromosomal  
 39 location distance between each pair of markers in the group is determined. The greatest absolute

chromosomal location distance between each pair of markers in the group is the length of the group of covering markers.

**A group of covering markers located on one chromosome can be ordered as a sequence of markers** starting with the marker closest to one end of the chromosome and going toward the other end of the chromosome. This is denoted for example as  $m_1, m_2, m_3, \dots, m_{N-2}, m_{N-1}, m_N$ , wherein  $N$  is the number of markers in the group. (The chromosomal location distance between  $m_1$  and  $m_N$  is greater than the chromosomal location distance between any other pair of markers in the group and this distance is the length of the group of markers.) **The chromosomal location distance between two successive markers in the group**, i.e. between  $m_R$  and  $m_{R+1}$ , is a **chromosomal intermarker distance**. (There are  $N-1$  chromosomal intermarker distances for a group of  $N$  covering markers.) **The average chromosomal intermarker distance** for a group is calculated by dividing the length of the group by  $(N-1)$ , wherein  $N$  is the number of covering markers in the group.

**The width of a CL-F region** is the largest frequency distance between any two CL-F points in the region.

**The length of a CL-F region** is the largest chromosomal location distance between any two CL-F points in the region.

**A segment-subrange pair** is the pair formed by pairing a segment of a chromosome and a subrange of the least common allele frequency range 0 to 0.5.

The **term segment-subrange** is used as a short version of the term segment-subrange pair. (A segment-subrange is a rectangular region on a CL-F map or a rectangular CL-F region, see below.)

**If one or more bi-allelic markers are said to be within(or in) a segment-subrange** then each of the markers is located on (or in) the chromosomal segment of the segment-subrange(pair) and each of the markers' least common allele frequencies is in the subrange of the segment-subrange(pair). (And each of the markers is located within the rectangular region defined by the segment-subrange on a CL-F map.)

Alternatively, **if a segment-subrange is said to contain one or more markers or to contain the location of one or more markers** then each of the markers is located on (or in) the chromosomal segment of the segment-subrange and each of the markers' least common allele frequencies is in the subrange of the segment-subrange. (And each of the markers is located within or is within the rectangular region on a CL-F map defined by the segment-subrange.)

**If one or more CL-F points are said to be within(or in) a segment-subrange** then each of the points is located within the rectangular region defined by the segment-subrange on a CL-F map or on the segment-subrange's borders.

**The length of a segment-subrange** is the length of the segment of the segment-subrange.

**The width of a segment-subrange** is the width of the subrange of the segment-subrange.

**The area of a segment-subrange** is the segment subrange's length multiplied by the segment subrange's width.

**If a CL-F region is said to comprise a segment-subrange**, then each point in the segment-subrange is in(or included in) the CL-F region.



1 If a CL-F region is said to comprise an area of greater than or equal to  $X$  multiplied by  $Y$ , then the  
2 CL-F region comprises one or more nonoverlapping segment-subranges, and the sum of the areas of  
3 the segment-subranges is greater than or equal to  $X$  multiplied by  $Y$ .

4 A CL-F matrix is a collection of segment-subranges, wherein each segment-subrange of the collection  
5 has the same width and the same length. Each segment-subrange in the collection (or the matrix) is a  
6 CL-F matrix cell. Any one CL-F matrix cell in a CL-F matrix shares two or more of the cell's borders  
7 with two or more other cells in the matrix. And all the cells in a CL-F matrix together form a single  
8 segment-subrange. A CL-F matrix is characterized by the length and the width of the cells in the matrix  
9 denoted by length  $\times$  width, or  $L_{MC} \times W_{MC}$ , wherein  $L_{MC}$  is the length of each cell in the matrix and  $W_{MC}$  is  
10 the width of each cell in the matrix. A CL-F matrix is also characterized by the number of rows of cells,  
11  $R_M$ , in the matrix. And a CL-F matrix is characterized by the number of columns of cells,  $C_M$ , in the  
12 matrix. There are two or more cells in a CL-F matrix. A CL-F matrix is also characterized by the point of  
13 origin of the matrix, denoted by  $(c_0, f_0)$ . The point of origin of a CL-F matrix is at any chromosomal  
14 location and  $c_0$  takes on any reasonable value in an entire species genome. The point of origin of a  
15 CL-F matrix is at any one value in the least common allele frequency range 0 to 0.5. (A CL-F matrix is  
16 similar to the squares of a chessboard or to equal rectangular floor tiles that are all oriented in the same  
17 direction and cover a rectangular floor. One corner of the matrix is the matrix's point of origin.)

18 The width of each cell of a particular CL-F matrix is any value greater than zero and less than 0.5.

19 The width of a cell is often denoted by  $W_{MC}$ .

20 Any length in chromosomal location distance units is chosen for the length of each cell of a particular  
21 CL-F matrix. The length of a cell is often denoted by  $L_{MC}$ .

22 The centerpoint of a CL-F matrix cell is in the center of the cell. The centerpoints of a CL-F matrix form  
23 a matrix centerpoint lattice. Each point of a matrix centerpoint lattice is separated by a CL-F distance  
24 of  $[0, W_{MC}]$  or  $[L_{MC}, 0]$  from two or more neighboring centerpoints.

25 If one or more bi-allelic markers are in (or within) the segment-subrange that is a CL-F matrix  
26 cell, then each of the markers is in or within the CL-F matrix cell.

27 If one or more CL-F points is in (or within) a CL-F matrix, then each of the points is in or within a cell  
28 of the matrix.

29 If a CL-F region comprises a CL-F matrix, then each point that is in the matrix is also in the region.

30 If a CL-F region is a CL-F matrix, then the region consists of the points that are in the matrix.

31 If two CL-F matrix cells share a common border, then the two CL-F matrix cells are in contact.

32 If two CL-F matrix cells share a common corner, then the two CL-F matrix cells are touching. (Two  
33 cells that are in contact are also touching.)

34 If a group of CL-F points is connected to within a CL-F distance  $[X, Y]$ , then for any two points in  
35 the group, denoted  $p_1$  and  $p_R$ , there is an ordered sequence of points in the group denoted  $p_1, p_2,$

36  $p_3, \dots, p_{R-2}, p_{R-1}, p_R, R$  being an integer greater than or equal to 2, wherein the CL-F distance between  
37 each point in the sequence and the next point in the sequence is less than or equal to  $[X, Y]$ . The

38 distance  $[X, Y]$  is the connecting distance. (Put in simple terms if a group of points is connected to  
39 within  $[X, Y]$ , then there is a path between each pair of points in the group, the path consisting of a

40 series of steps, wherein each step in the path is a movement between two points in the group that are

separated by a CL-F distance of less than or equal to  $[X, Y]$ . A simple group of points connected to within a CL-F distance of  $[X, Y]$  is a group of three points, wherein each point in the group is within a CL-F distance of less than or equal to  $[X, Y]$  of another point in the group. The concept of connectivity introduced here is similar to the basic concept of connectivity in mathematical graph theory.)

If a group of  $N$  markers is connected to within a CL-F distance  $[X, Y]$ , wherein  $N$  is an integer, then each of the markers is located at one point of a group of  $N$  points, the group of  $N$  points being connected to within a CL-F distance  $[X, Y]$ .

If two bi-allelic markers are said to be in extreme positive disequilibrium then  $d$  is approximately equal to  $d_{\max}$  for the two markers, which for the purposes of this definition are designated marker  $M$  with least common allele  $A$  and marker  $m$  with least common allele  $B$ . Wherein according to standard usage, the disequilibrium coefficient ( $d$ ) is defined by the equation  $d = f(AB) - f(A)f(B)$  where  $f(A)$  and  $f(B)$  are defined as the population frequencies of alleles  $A$  and  $B$ , respectively, and  $f(AB)$  is the population frequency of the  $AB$  haplotype. And  $d_{\max}$  is defined as the maximum possible positive value of  $d$  assuming the allele frequencies of  $A$  and  $B$  are  $f(A)$  and  $f(B)$ , and thus  $d_{\max} = \min\{f(A), f(B)\}$  where  $q$  is the lesser of  $f(A)$  and  $f(B)$ . (In this application  $d$  is used to represent the disequilibrium coefficient; the symbol  $\delta$  is often used in scientific papers to represent the disequilibrium coefficient.)

If a pair of markers is said to be in extreme positive disequilibrium, then the two markers of the pair are in extreme positive disequilibrium.

If a pair of bi-allelic markers is said to be redundant within distance  $D$  then the two markers of the pair are in extreme positive disequilibrium and the two markers are located on the same chromosome and the two markers are located within a CL-F distance  $D$  of each other on a CL-F map, wherein  $D$  is a specified distance and  $D$  has two components, a chromosomal location distance component  $D_{CL}$  and a frequency distance component,  $D_F$ ;  $D = [D_{CL}, D_F]$ .

An allele equivalent (AE) is a group of one or more "haplotype values" of one or more polymorphisms of the same type, either markers or genes. (For the purposes of this application a haplotype value of one polymorphism is equivalent to an allele value at the one polymorphism.) The group of haplotype values is then analyzed as if the group is a single allele at a bi-allelic polymorphism; the group of haplotype values acts as a single allele at a bi-allelic polymorphism; the collection of the one or more polymorphisms upon which the haplotype values are based acts as a bi-allelic polymorphism; the collection of one or more polymorphisms forms a bi-allelic polymorphism equivalent (PE) that acts as a bi-allelic polymorphism; the polymorphism equivalent has (or possesses) the allele equivalent. The allele equivalent belongs to the polymorphism equivalent. In this application, each polymorphism equivalent is a bi-allelic marker equivalent (BME) or a bi-allelic gene equivalent (BGE).

A bi-allelic marker equivalent (BME) is one or more markers and a grouping of the haplotype values of the one or more markers into two groups (e.g. group I and group II) (For the purposes of this application a "haplotype value" of one marker is equivalent to an allele at the one marker). The one or more markers and the two groups of haplotype values of the one or more markers are then analyzed as if the one or more markers are a single bi-allelic marker with alleles I and II. Each group of the groups I and II is an allele equivalent. For example, a multi-allelic microsatellite marker has its multiple alleles grouped into two groups and the microsatellite marker and these two groups of alleles then act

equivalent to a bi-allelic marker and are analyzed as if the microsatellite marker with the two groups is bi-allelic (for an example of this see McGinnis, Ewens & Spielman, Genetic Epidemiology 1995 ; 12(6) : 637-40, which is incorporated herein by reference)

Also for example, two or more multi-allelic markers have their haplotypes separated into two groups of haplotypes and the multi-allelic markers with their two groups of haplotypes are analyzed as if they were a single bi-allelic marker.

For example bi-allelic marker A has alleles a and a\* and bi-allelic marker B has alleles b and b\*. Then the four haplotypes ab, ab\*, a\*b\* and a\*b are grouped into two groups, for example group I: ab and a\*b\* and group II : ab\* and a\*b. Then a BME formed by markers A and B takes on values of group I (or I) for haplotypes ab or a\*b\* or group II (or II) for the haplotypes ab\* or a\*b ; and the two markers and the two group values( I and II) are analyzed as though they form a single bi-allelic marker(the BME). The same type of reasoning and procedure is extended to 3 or more bi-allelic markers, 3 or more bi-allelic marker equivalents or 2 or more multi-allelic markers.

(Logically, of course, the genotype at a BME for an individual is determined by knowing the two haplotype values at the one or more markers that form the BME for each of the individual's two homologous chromosomes that carry the one or more markers. The genotype is then determined by classifying each haplotype as belonging to group I or group II or the equivalent thereof. The three possible genotype values at the BME are I / I, I / II, and II / II or the equivalent thereof.)

Similarly, a **bi-allelic gene equivalent (BGE)** is one or more genes and a grouping of all the haplotype values of the one or more genes into two groups (e.g. group I and group II).

For the purposes of the description and claims, **the chromosomal location of a polymorphism equivalent** is at any point on the smallest chromosomal segment that contains the one or more polymorphisms that form the polymorphism equivalent(PE).

**The allele frequency of an allele equivalent (AE)** is determined as follows. An allele equivalent (AE) is a group of haplotype values of one or more polymorphisms. The frequency of the allele equivalent is the sum of the frequencies of the haplotype values in the group that makes up the allele equivalent.

For the purposes of the application, description, claims and definitions the term **true allele** is used to distinguish an allele according to standard usage (i.e. at a single polymorphism) from an allele equivalent (AE).

**The least common allele frequency of a bi-allelic polymorphism equivalent (BPE)** is determined as follows. Each of the two groups( I and II) of the haplotype values of the one or more polymorphisms which form the BPE is assigned a frequency. The frequency of I is the sum of the frequencies of the haplotype values in group I. And the frequency of II is the sum of the frequencies of the haplotype values in group II. The least of the frequency of I and the frequency of II is the least common allele frequency of the BPE. If the frequency of I and the frequency of II are equal, then the least common allele frequency of the BPE is the frequency of I or the frequency of II.

For the purposes of the description and claims, **the chromosomal location of a bi-allelic marker equivalent (BME)** is at any point on the smallest chromosomal segment which contains the one or more markers which form the BME.

1 **The chromosomal location distance from a BME to a CL-F point** on a CL-F map is the shortest  
2 chromosomal location distance from the CL-F point to any one of the one or more markers which form  
3 the BME.

4 **The least common allele frequency of a bi-allelic marker equivalent (BME)** is determined as  
5 follows. Each of the two groups( I and II) of the haplotype values of the one or more markers which form  
6 the BME is assigned a frequency. The frequency of I is the sum of the frequencies of the haplotype  
7 values in group I. And the frequency of II is the sum of the frequencies of the haplotype values in group  
8 II. The least of the frequency of I and the frequency of II is the least common allele frequency of the  
9 BME. If the frequency of I and the frequency of II are equal, then the least common allele frequency of  
10 the BME is the frequency of I or the frequency of II.

11 **The frequency distance from a BME to a CL-F point** on a CL-F map is the absolute difference  
12 between the least common allele frequency of the BME and the least common allele frequency  
13 coordinate of the CL-F point.

14 ( If a CL-F point on a CL-F map is covered by one or more BMEs to within a distance  $\delta$ , wherein  $\delta = [\delta_{CL}$   
15  $, \delta_F]$ , then the CL-F distance from each of the one or more BMEs to the CL-F point is less than or equal  
16 to  $\delta$ . And the chromosomal location distance from one of the markers which form each BME to the CL-F  
17 point is less than or equal to  $\delta_{CL}$ . And the frequency distance from each of the one or more BMEs to the  
18 CL-F point is less than or equal to  $\delta_F$ .)

19 **A bi-allelic marker equivalent is in(or within) each CL-F matrix cell that contains the**  
20 **chromosomal location of the bi-allelic marker equivalent (BME).** (Since the chromosomal location  
21 of a bi-allelic marker equivalent (BME) is at any point on the smallest chromosomal segment which  
22 contains the one or more markers which form the BME, in some cases, a bi-allelic marker equivalent is  
23 in more than one CL-F matrix cell.)

24 For the purposes of the application, the term **true bi-allelic marker** is used to distinguish a bi-allelic  
25 marker with two alleles according to usual usage (i.e. at a single polymorphism) from a bi-allelic marker  
26 equivalent(BME). A true bi-allelic marker is not a bi-allelic marker equivalent (BME). The term **true bi-**  
27 **allelic polymorphism** is used to distinguish a bi-allelic polymorphism with two alleles according to  
28 usual usage from a bi-allelic polymorphism equivalent(BPE).

29 The term **true allele** of a true bi-allelic marker means an allele of a true bi-allelic marker.

30 **A polymorphism(marker or gene) which is exactly bi-allelic** has exactly two alleles and the sum of  
31 the frequency of each of the two alleles is 1; for example if the two alleles are A and B, then  $f(A) + f(B)$   
32  $= 1$ . A polymorphism that is exactly bi-allelic is a true bi-allelic polymorphism with exactly two true  
33 alleles or a bi-allelic polymorphism equivalent (BPE) with exactly two allele equivalents.

34 **A polymorphism(marker or gene) which is approximately bi-allelic** has three or more alleles. And  
35 the polymorphism has a first allele and a second allele; and the sum of the frequency of the first allele  
36 and the frequency of the second allele is approximately 1. And the frequency of the first allele and the  
37 frequency of the second allele is much greater than the sum of the allele frequencies of all the alleles of  
38 the polymorphism that are not the first or the second alleles. For the versions of the invention for bi-  
39 allelic polymorphisms (bi-allelic markers and bi-allelic genes) described herein, a polymorphism which  
40 is approximately bi-allelic is analyzed as if the polymorphism has only two alleles, the first allele and the



1 **An oligonucleotide** is either a single or double stranded oligonucleotide. The length of an  
2 oligonucleotide ranges from a few bases or base pairs to approximately any number of bases or base  
3 pairs in the DNA sequence of any allele.

4 **An oligonucleotide, either single or double stranded, is complementary** to an allele if the DNA  
5 sequence of each strand of the oligonucleotide is exactly or approximately complementary to all or part  
6 of the DNA sequence of one of the DNA strands of the allele and the oligonucleotide has utility in  
7 identifying the allele by a hybridization reaction or equivalent thereof similar to as described below  
8 under oligonucleotide technology.

9 **An allele is identified by a hybridization reaction with an oligonucleotide that is complementary**  
10 **to the allele.** In this application **there are two types of oligonucleotides that are complementary**  
11 **an allele.** The two types of oligonucleotides complementary to an allele are identified as type(1) or  
12 type(2).

13 A type (1) complementary oligonucleotide is complementary to the part of an allele's DNA sequence  
14 that actually contains the allele's polymorphic site; and the type(1) complementary oligonucleotide has  
15 utility to identify the allele by means of a hybridization reaction of the oligonucleotide to the part of the  
16 allele's DNA sequence that actually contains the allele's polymorphic site. A hybridization reaction of a  
17 type(1) oligonucleotide to the part of an allele's DNA sequence that actually contains the allele's  
18 polymorphic site is a type (1) hybridization reaction.

19 A type (2) complementary oligonucleotide is complementary to an allele at a DNA sequence that flanks  
20 (but does not contain) the allele's polymorphic site; and the type (2) complementary oligonucleotide has  
21 utility to identify the allele by means of a hybridization reaction wherein the oligonucleotide hybridizes to  
22 the allele at a DNA sequence that flanks (but does not contain) the allele's polymorphic site and  
23 identification of the allele is subsequently achieved by extension of the oligonucleotide (and possibly  
24 one or more other type(2)complementary oligonucleotides) across the polymorphic site with a DNA  
25 polymerase such as occurs, for example, in a standard PCR (polymerase chain reaction). A  
26 hybridization reaction of a type(2) oligonucleotide to an allele at a DNA sequence that flanks (but does  
27 not contain) the allele's polymorphic site is a type (2) hybridization reaction.

28 Each version of **oligonucleotide technology** is a means to test for the presence (or absence) of each  
29 of one or more true alleles of a group of true alleles in an individual's chromosomal DNA. The presence  
30 or absence of any one true allele in the group is tested for by means of a type (1) or type (2)  
31 hybridization reaction (or equivalent) with an oligonucleotide that is complementary(type(1) or type(2))  
32 to the true allele. Put another way, the presence or absence of each true allele in the group is tested for  
33 by means of a type(1) or type(2)hybridization reaction (or equivalent) with an oligonucleotide that is  
34 complementary to each true allele in the group. There are many versions of oligonucleotide technology,  
35 some of these versions are described in more detail below. (In this application, the term "chromosomal  
36 DNA" includes chromosomal DNA obtained directly from an individual as well as DNA obtained as  
37 amplification products using PCR and chromosomal DNA obtained directly from an individual.  
38

**A physico-chemical signal** is any physical (including chemical) signal which is detected by human senses or by apparatus. A physico-chemical signal includes, but is not limited to, (1) an electrical signal such as is generated when oligonucleotides that are attached to a silicon chip hybridize with complementary alleles, (2) a visual or optical signal such as is generated when oligonucleotides attached to a glass slide hybridize with complementary alleles, (3) a signal (such as a dye color) generated by the products of a PCR (polymerase chain reaction) such as when oligonucleotides that are used as primers for PCR reactions hybridize with complementary alleles.

**The collection of true alleles of a group of one or more bi-allelic markers** is defined as consisting of each true allele of each true marker in the group and each true allele of each haplotype that forms each allele equivalent of each marker equivalent in the group.

**If a set of oligonucleotides is said to be complementary to a group of one or more bi-allelic markers**, then each oligonucleotide in the set is type(1) or type(2) complementary to at least one of the true alleles in the collection of true alleles of the group of one or more markers; and there is an oligonucleotide in the set that is type(1) or type(2) complementary to each true allele in the collection of true alleles of the group of one or more markers.

**Sample allele frequency data for a marker and a sample** is obtained by pooling DNA specimens from individuals of the sample into one or more DNA pools. An allele frequency for each of the marker's alleles is obtained for each DNA pool. In the case of a bi-allelic marker, determining the sample allele frequency for one allele essentially determines the sample allele frequency for the other allele. (For example, in some association based linkage studies, each DNA pool contains DNA from individuals of the sample with the same or similar phenotype status.) (It is also possible to obtain sample allele frequency for a marker and a sample by calculation using genotype data at the marker for each individual in the sample.)

**Genotype data/sample allele frequency data** for a marker and a sample is (1) genotype data at the marker for each individual of the sample, or (2) a combination of genotype data at the marker for one or more individuals in the sample and sample allele frequency data for the marker for the sample, or (3) sample allele frequency data for the marker for the sample. In the case of genotype data, DNA specimens from individuals are tested individually to determine genotype. In the case of sample allele frequency data DNA specimens from individuals are pooled, or sample allele frequency is calculated using genotype data for each individual in the sample.

### Description

For the versions of the invention described herein and the claims, **a bi-allelic genetic characteristic gene or a bi-allelic gene** is a gene which is exactly bi-allelic or a gene which is approximately bi-allelic. For the versions of the invention described herein and the claims, **a bi-allelic genetic characteristic gene or a bi-allelic gene** is a gene which is a true bi-allelic gene or a bi-allelic gene equivalent (BGE). A bi-allelic gene equivalent is exactly bi-allelic or approximately bi-allelic. A true bi-allelic gene is exactly bi-allelic or approximately bi-allelic.

For the versions of the invention described herein and the claims, a **bi-allelic marker** or a **bi-allelic covering marker** is a marker which is exactly bi-allelic or a marker which is approximately bi-allelic. Each marker that is exactly bi-allelic is a true bi-allelic marker or a bi-allelic marker equivalent. And each marker that is approximately bi-allelic is a true bi-allelic marker or a bi-allelic marker equivalent (BME).

**Process #1**, A process for identifying one or more bi-allelic markers linked to a bi-allelic genetic characteristic gene in a species of creatures, comprising the steps of :

- a) choosing two or more bi-allelic covering markers so that a CL-F region is systematically covered by the two or more covering markers;
- b) choosing a statistical linkage test based on allelic association for each covering marker;
- c) choosing a sample of individuals for each covering marker ;
- d) obtaining genotype data/sample allele frequency data for each covering marker and the sample chosen for each covering marker, and obtaining phenotype status data for the genetic characteristic for each individual in the sample chosen for each covering marker;
- e) calculating evidence for linkage between each covering marker and the gene using the statistical linkage test based on allelic association chosen for each covering marker and the genotype data/sample allele frequency data for each covering marker and using the phenotype status data for the genetic characteristic for each individual in the sample chosen for each covering marker obtained in d); and
- f) identifying those covering markers as linked to the genetic characteristic gene which show evidence for linkage based on the calculations of step e.

The following is a more detailed description of process #1.

**Process #1**, A process for identifying one or more bi-allelic markers linked to a bi-allelic genetic characteristic gene in a species of creatures comprising the steps of :

- a) choosing two or more bi-allelic covering markers so that a CL-F region is systematically covered by the two or more covering markers; Any method of systematically covering the CL-F region is acceptable. In this application, the systematic covering of a CL-F region in versions of the invention is described mathematically as the covering of a CL-F region, wherein the CL-F region is N covered to within a CL-F distance  $\delta$  by two or more bi-allelic covering markers. For further details



1 regarding this step, see Detailed Description of the Systematic Covering of a CL-F Region Used In  
2 Versions of the Invention below.

3  
4 **b)choosing a statistical linkage test based on allelic association for each covering marker ;** The  
5 statistical linkage test based on allelic association chosen for any one particular covering marker is any  
6 statistical linkage test based on allelic association as defined in the definitions section. Statistical  
7 linkage tests based on allelic association are described in the genetics and population genetics  
8 literature and are known to those of ordinary skill in the art. Some examples of a statistical linkage test  
9 based on allelic association are the TDT, Haplotype Relative Risk Method(HRR) and Allele Frequency  
10 Comparison In Disease Cases Versus Unrelated Controls .It is possible for different statistical linkage  
11 tests based on allelic association to be chosen for different covering markers. For purposes of technical  
12 convenience, the same statistical linkage test based on allelic association is preferably chosen for each  
13 covering marker.

14  
15 **c)choosing a sample of individuals from the species for each covering marker ;** For the process  
16 to be workable, the sample chosen for any one covering marker must be suitable for the statistical  
17 linkage test of b) above chosen for the covering marker. Knowledge of a suitable sample for the  
18 statistical linkage test chosen in b) above for the covering marker is within the understanding of a  
19 person skilled in the art. For purposes of technical convenience, the same sample of individuals is  
20 preferably chosen for each covering marker.

21  
22 **d)obtaining genotype data/sample allele frequency data for each covering marker and the**  
23 **sample chosen for each covering marker, and obtaining phenotype status data for the genetic**  
24 **characteristic for each individual in the sample chosen for each covering marker;**  
25 Sample allele frequency data for any one covering marker for the sample chosen for the covering  
26 marker is obtained by pooling DNA from individuals of the sample into one or more DNA pools. It is also  
27 possible to obtain sample allele frequency data for any one covering marker by calculation using  
28 genotype data at the marker for each individual in the sample. Each DNA pool contains DNA from  
29 individuals of the sample with the same or similar phenotype status. An allele frequency for each of the  
30 marker's alleles is obtained for each pool. Genotype data/sample allele frequency data for any one  
31 covering marker is (1)genotype data at the covering marker for each individual in the sample chosen for  
32 the covering marker, or (2)a combination of genotype data at the covering marker for one or more  
33 individuals in the sample chosen for the covering marker and sample allele frequency data for the  
34 covering marker for the sample chosen for the covering marker, or (3)sample allele frequency data for  
35 the covering marker for the sample chosen for the covering marker. The genotype data/sample allele  
36 frequency data for any one covering marker must be suitable for the statistical linkage test based on  
37 allelic association chosen for the covering marker in step b). It is possible to choose different types of  
38 genotype data/sample allele frequency data for each covering marker. For purposes of technical  
39 convenience, the same type of genotype data/sample allele frequency data (1), (2), or (3) is chosen for

each covering marker. Some examples of ways to practice this step is the use of technology cited under Oligonucleotide Technology (below) or mass spectrometry (such as MALDITOF) <sup>i</sup>

**e)calculating evidence for linkage between each covering marker and the gene using the statistical linkage test based on allelic association chosen for each covering marker and the genotype data/sample allele frequency data for each covering marker and using the phenotype status data for the genetic characteristic for each individual in the sample chosen for each covering marker obtained in d); and**

**f)identifying those covering markers as linked to the gene which show evidence for linkage based on the calculations of step e.**

The meanings of steps d, e and f are within the understanding of those of ordinary skill in the art. Fine points of using a statistical linkage test based on allelic association as a measure of evidence for linkage are known to those in the art. <sup>ii</sup>

Process #1 described above is equivalent to localizing a genetic characteristic gene to a particular chromosomal location (i.e. a sub-region of a particular chromosome.) This is because markers which are linked to a gene are also physically close to the gene in terms of physical (chromosomal) location. To locate a gene causing the genetic characteristic of Process #1, the gene is localized to the approximate chromosomal location of one or more covering markers which are identified as showing evidence for linkage in step f).

**Process#1A** It is also possible to use Process #1 to localize a genetic characteristic gene to an approximate CL-F location(chromosomal location-least common allele frequency location). Such a process is expressed as follows:

**Process#1A : A process for localizing a bi-allelic genetic characteristic gene in a species of creatures to a chromosomal location-least common allele frequency (CL-F) location, comprising the steps a), b), c), d) and e) of Process #1 and further comprising the step of:**

**f)localizing the gene to the chromosomal location-least common allele frequency (CL-F) location of one or more markers that show evidence for linkage based on the calculations of step e).**

It is the teaching of this application that the strength of evidence for linkage increases as markers that are in linkage disequilibrium with a gene become close to the gene on a CL-F map. It is possible for step f) to be done by an individual plotting data by hand and examining the data. It is also possible for software to perform step f). It is possible for this step to include using the dependence of quantitative evidence for linkage of step e) on CL-F location. For example, if quantitative evidence for linkage calculated in step e) (of process #1 or #1A) is represented in the z dimension of a typical three-dimensional x-y-z plot, wherein the x and y dimensions are chromosomal location and least common allele frequency respectively, then it is possible to conceptualize evidence for linkage as occurring in a "hump" (or "humps" )in the z dimension. And it is possible to use the evidence for linkage calculated in step e) of (process #1 or #1A) to find the CL-F location (in the x-y plane) of the peak(s) of a "hump(s)",

- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- 10
- 11
- 12
- 13
- 14
- 15
- 16
- 17
- 18
- 19
- 20
- 21
- 22
- 23
- 24
- 25
- 26
- 27
- 28
- 29
- 30
- 31
- 32
- 33
- 34
- 35
- 36
- 37
- 38
- 39
- 40
- 41

5  
6  
7

## 8

9  
10  
11  
12  
13  
14  
15  
16

## 17

18  
19  
20  
21  
22  
23  
24

25  
26  
27  
28  
29  
30  
31  
32

33  
34

35  
36  
37  
38  
39

40  
41

polymorphisms in any population of the species. It is preferable in terms of increasing the power of a version of the invention for linkage studies that the chromosomal location component of  $\delta$  be less than about the greatest chromosomal length, computed by any method, for which linkage disequilibrium has been observed between any polymorphisms in any population of the species. In general, the smaller the chromosomal location component of  $\delta$ , the greater the power of a version of the invention for linkage studies.

It is possible for the frequency distance component of  $\delta$  to be as great as about 0.2. ( Depending on the penetrance ratio ( $r$ ) or the disequilibrium between marker and gene, it is also possible for the frequency distance component of  $\delta$  to be greater than 0.2 under some conditions as evidenced by Table 2 under Theory of Operation. So it is also possible for the frequency distance component of  $\delta$  to be as great as about 0.25 or higher.) It is preferable in terms of increasing the power of a version of the invention for linkage studies that the frequency distance component of  $\delta$  to be less than about 0.2. In general, the smaller the frequency distance component of  $\delta$ , the greater the power of a version of the invention for linkage studies.

Linkage disequilibrium has been observed between polymorphisms separated by 10 to 12 cM in some homogeneous human populations. Therefore, it is possible for the chromosomal location distance component of  $\delta$  to be as large as about 10 to 12 cM, about 10 to 12 million bp, or the equivalent thereof for homogeneous human populations. It is preferable in terms of increasing the power of a version of the invention for linkage studies in human populations that  $\delta$  is less than or equal to about [ 1 million bp, 0.15] or the equivalent thereof. It is more preferable in terms of increasing the power of a version of the invention for linkage studies in human populations that  $\delta$  is less than or equal to about [ 250,000 bp, 0.1] or the equivalent thereof.

In general, the smaller the magnitude of  $\delta$  is in terms of either frequency distance, chromosomal location distance, or both, the greater the power of a version of the invention for linkage studies. In general, the greater  $N$  is, the greater the power of a version of the invention for linkage studies.

Because the greater  $N$  is, the greater the chance that linkage is detected between one or more covering markers and a gene or genes. The largest that  $N$  is chosen is limited by the number of known markers in the neighborhood of the CL-F region and also by the distribution of the known markers.

In general, the larger the CL-F region which is  $N$  covered, the greater the power of a version of the invention for linkage studies, because a larger region is scanned (covered). Less dense coverings wherein  $N$  is small, and the magnitude of  $\delta$  is large also have technical and economic advantages for certain situations.

#### **Specific types of CL-F regions that are $N$ covered**

Specific types of CL-F regions that are  $N$  covered are useful. For example, a rectangular CL-F region, a segment-subrange, that is  $N$  covered is used in an association based linkage study to test for the presence of a trait causing bi-allelic gene located within the segment-subrange. In the case in which a group of points is  $N$  covered to within a CL-F distance  $[x,y]$  and the group of points is connected to within a CL-F distance of  $[2x,2y]$  or less, then a path connected CL-F region is  $N$  covered to within the CL-F distance  $[x,y]$ .

1 A CL-F matrix is a device to illustrate and describe the systematic nature of special cases of CL-F  
2 regions that are N covered. In the case in which there are N or more markers within each cell of a CL-F  
3 matrix, then each point within the matrix is N covered to within the CL-F distance  $[L_{CM}, W_{CM}]$ , wherein  
4  $L_{CM}$  is the length of a matrix cell and  $W_{CM}$  is the width of a matrix cell. A choice of covering markers so  
5 that approximately the same number of covering markers are in each cell of a CL-F matrix has utility in  
6 that approximately the same amount of effort is expended on each subregion (cell) of the CL-F region  
7 defined by the matrix in a linkage study using the covering markers. If the centerpoints of a CL-F matrix  
8 (a matrix centerpoint lattice) are each N covered by a group of covering markers to within a CL-F  
9 distance  $[x,y]$ , then each point in the matrix is N covered to within the CL-F distance  $[2x,2y]$ . A CL-F  
10 matrix can be used as a device to help distinguish versions of the invention from prior art (to the extent  
11 that there is prior art).

12 A requirement that the CL-F region that is N covered to within a certain CL-F distance comprise a  
13 certain minimum area or segment-subrange with a certain minimum area is a special case of CL-F  
14 regions that are N covered to within the certain CL-F distance. A requirement that the CL-F region that  
15 is N covered to within a certain CL-F distance has a certain length or width is a special case of CL-F  
16 regions that are N covered to within the certain CL-F distance. Each of these requirements is also a  
17 device that can be used to help distinguish versions of the invention from prior art.

18 **A Note on the Equivalence of Working With Individual Alleles of Markers to Perform Two-**  
19 **dimensional Linkage Studies and the CL-F approach using bi-allelic markers**

20 It is possible to conceptualize performing two-dimensional linkage studies wherein individual marker  
21 alleles are used to cover a two-dimensional space, rather than individual bi-allelic markers. Any  
22 individual marker allele is assigned a two-dimensional location consisting of the chromosomal location  
23 of the marker and the allele frequency of the marker allele. Two-dimensional chromosomal location-  
24 allele frequency spaces(or regions) are systematically covered by sets of covering alleles. Each  
25 individual covering allele is tested for association with a genetic characteristic. Versions of inventions  
26 using systematic chromosomal location-allele frequency(CL-AF) region coverings that are similar to  
27 versions of the invention in this application are possible. Indeed these types of inventions have been  
28 described in U.S.Provisional Patent Applications previously filed by the inventor.

29 However, such a conceptual framework and the resulting inventions are equivalent to the CL-F versions  
30 approach used in this application. This is because any marker allele, A, that is used as a covering allele  
31 can be made to be an allele equivalent of a bi-allelic marker equivalent(BME). So that a BME with allele  
32 equivalents A and nonA is a bi-allelic marker with allele A. Therefore, any set of covering alleles that  
33 systematically cover a two-dimensional CL-AF region is equivalent to a set of BMEs that systematically  
34 cover an equivalent CL-F region. Testing each covering allele for association with a genetic  
35 characteristic is exactly equivalent to testing each BME of a set of BMEs for evidence of linkage to a  
36 gene using a statistical linkage test based on allelic association. Even testing for the presence or  
37 absence of a covering allele in the chromosomal DNA of an individual is equivalent to genotyping the  
38 individual at a BME. And determining a sample allele frequency for a covering allele, is equivalent to  
39 determining the sample allele frequencies for a BME.

**Example 1 of Process #1 is used for identifying markers linked to a disease gene.**

Example 1 A process for identifying bi-allelic markers linked to a bi-allelic disease gene in human beings, comprising the steps of :

- a) choosing two or more bi-allelic covering markers so that a CL-F region is N covered to within a CL-F distance [ 250,000 bp, 0.1] or the equivalent thereof by the covering markers, wherein N is an integer number greater than or equal to 2 ;
- b) choosing the same statistical linkage test based on allelic association for each covering marker;
- c) choosing the same sample of individual human beings for each covering marker;
- d) obtaining genotype data at each covering marker for each individual in the sample and obtaining phenotype status data for the disease for each individual in the sample ;
- e) calculating evidence for linkage between each covering marker and the gene using the test chosen in step b) and the genotype data at each covering marker and the using the phenotype status data for the disease for each individual in the sample ; and
- f) identifying those covering markers as linked to the gene which show evidence for linkage based on the calculations of step e.

**Apparatus Versions**

**General step by step descriptions of individual apparatus versions are given below.**

**Apparatus #1, an apparatus to practice process #1.**

**Apparatus #1**, An apparatus for identifying bi-allelic markers linked to a bi-allelic genetic characteristic gene in a species of creatures, comprising :

- a) means for choosing two or more bi-allelic covering markers so that a CL-F region is systematically covered by the two or more covering markers;
- b) means for choosing a statistical linkage test based on allelic association for each covering marker;
- c) means for choosing a sample of individuals for each covering marker ;

d) means for obtaining genotype data/sample allele frequency data for each covering marker and the sample chosen for each covering marker, and for obtaining phenotype status data for the genetic characteristic for each individual in the sample chosen for each covering marker;

e) means for calculating evidence for linkage between each covering marker and the gene using the statistical linkage test based on allelic association chosen for each covering marker and the genotype data/sample allele frequency data for each covering marker and using the phenotype status data for the genetic characteristic for each individual in the sample chosen for each covering marker obtained in d); and

f) means for identifying those covering markers as linked to the gene which show evidence for linkage based on the calculations by means e).

**More detailed description of Apparatus #1:** Apparatus #1 is an apparatus to practice process #1. More details of the description of apparatus #1 are found under the description of Process #1 above. Any one of the means labeled a), b), c), d), e) or f) of apparatus #1 includes any means for automating or partially automating a step as step a), b), c), d), e) or f) respectively of process #1. An example of any one of the means in this paragraph labeled a), b), c), d), e), or f) is means comprising an appropriately programmed, suitable computer, the computer being supplied with proper data and instructions.

The means labeled d) of apparatus #1 for obtaining genotype data/ sample allele frequency data for each covering marker for the sample chosen for each covering marker includes any automated or partially automated means to obtain genotype data/ sample allele frequency data. An example of means to obtain genotype data/ sample allele frequency data is means using mass spectrometry.<sup>1</sup> Means to obtain genotype data/ sample allele frequency data that is automated or partially automated includes means comprising Oligonucleotide Technology described below.

**Apparatus #1A, an apparatus to practice process #1A.**

**Apparatus#1A :** An apparatus for localizing a bi-allelic genetic characteristic gene in a species of creatures to a chromosomal location-least common allele frequency (CL-F) region, comprising the means a), b), c), d) and e) of Apparatus #1 and further comprising the means of: f) means for localizing the gene to the approximate chromosomal location-least common allele frequency region (CL-F) of one or more markers that show evidence for linkage based on the calculations of means e).





versions of the invention is described mathematically as the covering of a CL-F region, wherein the CL-F region is N covered to within a CL-F distance  $\delta$  by two or more bi-allelic covering markers. For further details regarding this, see Detailed Description of the Systematic Covering of a CL-F Region Used In Versions of the Invention above.

**An example of ApparatusGd/Safd#1Genotype data/Sample allele frequency data apparatus, a sample allele frequency apparatus:**

Example 1 of ApparatusGd/Safd#1: An apparatus for obtaining genotype data/sample allele frequency data for each bi-allelic marker of a group of two or more bi-allelic covering markers in the chromosomal DNA of one or more individuals of a sample, wherein the genotype data/sample allele frequency data is sample allele frequency data, comprising:

a) means for determining information on the presence or absence of each allele of each bi-allelic marker of a group of two or more bi-allelic covering markers in the chromosomal DNA from one or more individuals of the sample, a CL-F region being N covered to within the CL-F distance [ 1.0 cM, 0.15] by the two or more bi-allelic covering markers, wherein N is an integer number greater than or equal to 1, and

b) means for transforming the information of step a) into sample allele frequency data for each marker of the group.

Example 2 of ApparatusGd/Safd#1: An apparatus for obtaining genotype data/sample allele frequency data for each bi-allelic marker of a group of two or more bi-allelic covering markers in the chromosomal DNA of an individual, wherein the genotype data/sample allele frequency data is genotype data, comprising:

a) means for determining information on the presence or absence of each allele of each bi-allelic marker of a group of two or more bi-allelic covering markers in the chromosomal DNA from an individual, a CL-F region being N covered to within the CL-F distance [12cM, 0.25] or the equivalent thereof by the two or more bi-allelic covering markers, wherein N is an integer number greater than or equal to 1; and

b) means for transforming the information of step a) into genotype data for each marker of the group.

( It should be noted that the following genotype apparatus is equivalent to Example 2 of ApparatusGd/Safd#1: Genotype Apparatus: An apparatus for genotyping an individual, comprising:

a) means to genotype an individual at two or more bi-allelic covering markers, a CL-F region being N covered to within the CL-F distance [12cM, 0.25] or the equivalent thereof by the two or more bi-allelic covering markers, wherein N is an integer number greater than or equal to 1. )

1 **Genotype data/Sample allele frequency data process**

2 A process to obtain genotype data/sample allele frequency data similar to the data of the step d) of  
3 process #1 has great utility in that it is used to provide genotype data /sample allele frequency data for  
4 the more powerful two-dimensional linkage studies introduced in this application.

5 **Description of the Genotype data/Sample allele frequency data process.**

6 **ProcessGd/Safd#1: Genotype data/Sample allele frequency data process: A process for**  
7 **obtaining genotype data/sample allele frequency data for each bi-allelic marker of a group of**  
8 **two or more bi-allelic covering markers in the chromosomal DNA of one or more individuals of a**  
9 **sample, comprising:**

- 10 **a) determining information on the presence or absence of each allele of each bi-allelic marker of**  
11 **a group of two or more bi-allelic covering markers in the chromosomal DNA of one or more**  
12 **individuals of the sample, a CL-F region being systematically covered by the two or more bi-**  
13 **allelic covering markers; and**
- 14 **b) transforming the information of step a) into genotype data/sample allele frequency data for**  
15 **each marker of the group.**

16 The CL-F region and covering markers are for a species and the one or more individuals are members  
17 of the species. Determining information on the presence or absence of each allele of each bi-allelic  
18 marker of the group in chromosomal DNA includes any method of determination. Determining  
19 information on the presence or absence of each allele of each bi-allelic marker of the group in  
20 chromosomal DNA includes methods comprising oligonucleotide technology by using a set of  
21 oligonucleotides that is complementary to the group as discussed below. Information on the presence  
22 or absence of each allele in the chromosomal DNA is obtained using a DNA specimen from each of one  
23 or more individuals of the sample or by using one or more DNA pools of DNA specimens from two or  
24 more individuals of the sample. Any process that obtains genotype data or sample allele frequency data  
25 (similar to the data of the step d) of process #1) by determining the presence or absence of each allele  
26 of each bi-allelic marker of the group in the chromosomal DNA of one or more individuals is an example  
27 of this version of the invention. Versions of this process also obtain a combination of genotype data and  
28 sample allele frequency data similar to the data of the step d) of process #1. The details of step b) will  
29 be clear to those of ordinary skill in the art.

30

31 Each bi-allelic covering marker is a true bi-allelic or BME. Determining the presence or absence of each  
32 allele of each bi-allelic marker in the group includes determining the presence or absence of each allele  
33 equivalent of each bi-allelic marker equivalent(BME) in the group. Any method of systematically  
34 covering the CL-F region is acceptable. In this application, the systematic covering of a CL-F region in  
35 versions of the invention is described mathematically as the covering of a CL-F region, wherein the CL-  
36 F region is N covered to within a CL-F distance  $\delta$  by two or more bi-allelic covering markers. For further

PCT/US 99/04376  
IPEA/US 24 MAY 2001

33

1 details regarding this, see Detailed Description of the Systematic Covering of a CL-F Region Used In  
2 Versions of the Invention above.

3 **An example of ProcessGd/Safd#1 Genotype data/Sample allele frequency data process, a**  
4 **genotype data process:**

5 Example 1 of ProcessGd/Safd#1: A process for obtaining genotype data/sample allele frequency data  
6 for each bi-allelic marker of a group of two or more bi-allelic covering markers in the chromosomal DNA  
7 of an individual, wherein the genotype data/sample allele frequency data is genotype data, comprising:

8 a) determining information on the presence or absence of each allele of each bi-allelic  
9 marker of a group of two or more bi-allelic covering markers in the chromosomal DNA from an  
10 individual, a CL-F region being N covered to within the CL-F distance [12cM, 0.25] or the equivalent  
11 thereof by the two or more bi-allelic covering markers; wherein N is an integer number greater than or  
12 equal to 1; and

13 b) transforming the information of step a) into genotype data for each marker of the group.

14 ( It should be noted that the following genotype process is equivalent to Example 1 of  
15 ProcessGd/Safd#1: Genotype Process: A process for genotyping an individual, comprising:

16 a) genotyping an individual at two or more bi-allelic covering markers, a CL-F region being N  
17 covered to within the CL-F distance [12cM, 0.25] or the equivalent thereof by the two or more bi-allelic  
18 covering markers, wherein N is an integer number greater than or equal to 1. )

20 **Oligonucleotide technology**

21 Each version of oligonucleotide technology is a means to sense the presence or absence of each of  
22 one or more true alleles of a group of true alleles in chromosomal DNA from one or more individuals by  
23 means of a hybridization reaction with an oligonucleotide that is complementary to each of the one or  
24 more true alleles (see definitions section). Thus versions of oligonucleotide technology are a means of  
25 genotyping one or more individuals. And, versions of oligonucleotide technology are a means of  
26 obtaining sample allele frequency data for one or more marker alleles for a sample of individuals using  
27 pooled DNA from the individuals in the sample.

28 In Some Versions of Oligonucleotide Technology for Genotyping or Obtaining Sample Allele Frequency  
29 Data, a Physico-chemical Signal is Generated when an Allele in Chromosomal DNA and a  
30 Complementary Oligonucleotide Hybridize

31 Some versions of oligonucleotide technology for genotyping or for obtaining sample allele frequency  
32 data use a sensor which includes one or more oligonucleotides which are complementary to an allele.  
33 When the sensor is exposed to chromosomal DNA from an individual who carries the allele, the  
34 oligonucleotides which are complementary to the allele hybridize with chromosomal DNA specimens of  
35 the allele. The hybridization generates a physico-chemical signal which indicates the presence of the

AMENDED SHEET

SCANNED # 14

PCT/US 99/04376

IPEA/US 24 MAY 2000

allele in the chromosomal DNA of the individual. The lack of the physico-chemical signal indicates no (or negligible) hybridization and that the allele is not present in the chromosomal DNA of an individual.

Examples of oligonucleotide technology for genotyping, obtaining sample allele frequency data or genotype data/sample allele frequency data

Companies like Affymetrix are using high density arrays of oligonucleotides attached to silicon chips or glass slides to genotype DNA from one individual at thousands of bi-allelic markers.<sup>i</sup> In some of these versions of oligonucleotide technology, the strength of hybridization of oligonucleotides that differ at only one base to DNA containing an SNP are compared to determine genotype.<sup>ii</sup> Another version of oligonucleotide technology uses oligonucleotides as PCR (Polymerase Chain Reaction) primers to obtain genotype data.<sup>iii</sup> Other examples of oligonucleotide technology and its uses to obtain genetic information are included in the articles cited in the endnotes.<sup>iv</sup> Versions of oligonucleotide technology obtain sample allele frequency data from pooled DNA or genotype data using oligonucleotides as PCR primers to obtain amplified reaction products that are detected by mass spectrometry. Another example of oligonucleotide technology is padlock probes.<sup>v</sup>

Other examples of oligonucleotide technology are minisequencing on DNA arrays, dynamic allele-specific hybridization, microplate array diagonal gel electrophoresis, pyrosequencing, oligonucleotide-specific ligation, the TaqMan system and immobilized padlock probes as presented at the First International Meeting on Single Nucleotide Polymorphism and Complex Genome Analysis.<sup>vi</sup>

Sets of Oligonucleotides for Genotyping at Bi-allelic Markers or Obtaining Sample Allele Frequency Data

*A set of oligonucleotides that is complementary (see definitions) to a group of one or more bi-allelic markers has utility to determine genotype data at each of the markers in the group, including groups with BMEs and approximately bi-allelic markers.*

Similarly, a set of oligonucleotides that is complementary to a group of bi-allelic markers has utility to obtain sample allele frequency data for each allele of each marker in the group.

*In both cases, obtaining genotype data or sample allele frequency data, the same principle is used: a set of oligonucleotides that is complementary to a group of bi-allelic markers has utility to determine the presence or absence of each allele of each marker in the group in chromosomal DNA.*

Using sets of oligonucleotides to obtain Genotype Data/Sample Allele Frequency Data for each marker of a group of bi-allelic markers, wherein the group of markers systematically cover a CL-F region

Genotype data/sample allele frequency data for each marker of a group of bi-allelic markers, wherein the group of bi-allelic markers systematically cover a CL-F region has great utility for use in the more powerful two-dimensional linkage studies introduced in this application. As described above under Oligonucleotide Technology, some sets of oligonucleotides have utility to determine genotype data at each bi-allelic marker of a group of one or more bi-allelic markers. Similarly, some sets of oligonucleotides have utility to obtain sample allele frequency data for each bi-allelic marker of a group of one or more bi-allelic markers. Therefore, the use of one or more copies of a set of oligonucleotides to obtain genotype data or sample allele frequency data for each bi-allelic marker of a group of one or

more bi-allelic covering markers, wherein the group of bi-allelic covering markers systematically cover a CL-F region has great utility.

A word to avoid confusion in terminology: in this application, a set of markers for use in genotyping is referred to as a set of oligonucleotides.

A set of oligonucleotides consisting of one or both strands of each allele of a group of one or more markers is a set of oligonucleotides that is complementary to the group of markers. (see definitions section) Such a set of oligonucleotides is in effect the group of markers themselves; and such a set of oligonucleotides has utility to determine genotype data at each marker in the group. So a group of markers (or set of markers) for use in obtaining genotype data or sample allele frequency data for each of the markers in the group is included in the descriptive phrase: "a set of oligonucleotides".

**Description of Use set#1 D:**

**Use set#1 D** The use of one or more copies of a set of oligonucleotides to determine genotype data/sample allele frequency data for each bi-allelic marker of a group of two or more bi-allelic covering markers for one or more individuals, wherein the group of covering markers systematically cover a CL-F region.

The CL-F region and covering markers are for a species and the one or more individuals are members of the species. An example of a set of oligonucleotides with utility to be used to determine genotype data/sample allele frequency data for each bi-allelic marker of a group of two or more bi-allelic covering markers is a set of oligonucleotides that is complementary to the group of markers. A set that is complementary to the group of markers is used to detect the presence or absence of each the alleles of the covering markers by means of a hybridization reaction as discussed under oligonucleotide technology. Thus a set that is complementary to the group of markers is used to determine genotype data/sample allele frequency data for each covering marker.

The use of one or more copies of a set of oligonucleotides to obtain genotype data or sample allele frequency data for each bi-allelic marker of a group of one or more bi-allelic covering markers, wherein the group of bi-allelic covering markers systematically cover a CL-F region are both examples of this version of the invention(Use Set#1D).

In this application, the systematic covering of a CL-F region in versions of the invention is described mathematically as the covering of a CL-F region, wherein the CL-F region is N covered to within a CL-F distance  $\delta$  by two or more bi-allelic covering markers. For further details regarding this, see Detailed Description of the Systematic Covering of a CL-F Region Used In Versions of the Invention above.

**Example 1S of Use set#1D:** The use in genotyping one or more individuals, of one or more copies of a set of oligonucleotides, the set of oligonucleotides being complementary to a group of two or more bi-allelic covering markers, a CL-F region being N covered by the covering markers to within a CL-F distance of about [ 250,000 bp, 0.1] or the equivalent thereof, wherein N is an integer greater than or equal to two.

**Composition of matter: Description of Comp set#1D:**

**Comp set#1D:** One or more copies of a set of oligonucleotides, the set of oligonucleotides being

1 **complementary to a group of two or more bi-allelic covering markers, wherein the group of**  
 2 **covering markers systematically cover a CL-F region.**

3 A set of oligonucleotides that is complementary to a group of two or more bi-allelic covering markers,  
 4 wherein the group of covering markers systematically cover a CL-F region has great utility for use in the  
 5 two-dimensional linkage study techniques introduced in this application. Such a set has utility in being  
 6 used to genotype individuals or obtain sample allele frequency data or genotype data/sample allele  
 7 frequency data as described above under Use set#1D. In this application, the systematic covering of a  
 8 CL-F region in versions of the invention is described mathematically as the covering of a CL-F region,  
 9 wherein the CL-F region is N covered to within a CL-F distance  $\delta$  by two or more bi-allelic covering  
 10 markers. For further details regarding this, see Detailed Description of the Systematic Covering of a CL-  
 11 F Region Used In Versions of the Invention above.

### 12 13 **Example 1Comp of Comp set#1D:**

14 **Example 1Comp:** One or more copies of a set of oligonucleotides, the set of oligonucleotides being  
 15 complementary to a group of two or more bi-allelic covering markers, a CL-F region being N covered by  
 16 the covering markers to within a CL-F distance of about [ 1cM, 0.2] or the equivalent thereof, wherein N  
 17 is an integer greater than or equal to one.

### 18 **Redundancy of Covering Markers**

19 Some versions of the invention make use of N coverings of CL-F regions by covering markers which  
 20 limit (possibly to zero) the number of pairs of covering markers which are redundant within CL-F  
 21 distance D,  $D = [D_{CL} \ D_F]$ , wherein D is less than or equal to about  $\delta$ , a CL-F covering distance. This  
 22 limits the number covering markers which are separated by a CL-F distance of less than or equal to D (if  
 23 the markers were placed on a CL-F map) which *will be in extreme positive disequilibrium with each*  
 24 *other*. This limitation is done by requiring that less than or equal to R pairs of covering markers are  
 25 redundant within distance D. Wherein R is an integer greater than or equal to 0 and less than or equal  
 26 to about  $N(N-1)/2$ . When R is chosen to be zero, no pair of covering markers is redundant within  
 27 distance D.

28 A preferable condition is that each bi-allelic covering marker within each small CL-F region (a small  
 29 segment-subrange of length about  $\delta_{CL}$  and width about  $\delta_F$  the distance components of the covering  
 30 distance  $\delta$  ) provides much new (i.e. non-redundant) information about linkage and association to any  
 31 nearby bi-allelic gene. Under these conditions, testing each bi-allelic covering marker in each small CL-  
 32 F region increases the likelihood of detecting linkage to a gene.

33 Limiting (including to zero) pairs of covering markers which are redundant within CL-F distance D (which  
 34 is less than or equal to a covering distance  $\delta$  ) approaches and achieves this preferable condition. This  
 35 limitation is not crucial to the functioning of a version of the invention, however, it has the advantage of  
 36 reducing excess effort and increasing efficiency.

### 37 **Polymorphism CL-F Display**

38 Polymorphism CL-F display apparatus display the chromosomal location, least common allele  
 39 frequency and identity of each polymorphism of one or more polymorphisms (markers or genes or both)

1 of one or more populations of one or more species on one or more two-dimensional graphs, each graph  
2 is similar to an x-y plot. The apparatus has utility including aiding in decisions regarding linkage studies  
3 and the interpretation of linkage study data.

4 The apparatus comprise means to display the chromosomal location, least common allele frequency  
5 and identity of each polymorphism of one or more polymorphisms (markers or genes or both) of one or  
6 more populations of one or more species on one or more two-dimensional graphs, each graph is similar  
7 to an x-y plot.

8 Each graph has two axes, one axis, the frequency axis, represents least common allele frequency and  
9 the alternate(or other) axis, the chromosomal location axis, represents chromosomal location. Each  
10 frequency axis of each graph is in units of population frequency. Each chromosomal location axis of  
11 each graph is in units of chromosomal location such as centimorgans, base pairs or the equivalent  
12 thereof.

13 The frequency axis of each graph spans the entire range 0 to 0.5 or a subrange of the range 0 to 0.5.

14 The chromosomal location axis of each graph spans the chromosomal locations on one or more  
15 segments of one or more chromosomes of a species, each of the one or more segments is a size from  
16 the equivalent of a base pair in length to the length of an entire chromosome (or the equivalent thereof).

17 Each point on each graph is directly opposite a value on the frequency axis of each graph. The value  
18 on the frequency axis directly opposite each point on each graph is the frequency coordinate of each  
19 point on each graph. Each point on each graph is directly opposite a value on the chromosomal location  
20 axis of each graph. The value on the chromosomal location axis directly opposite each point on each  
21 graph is the chromosomal location coordinate of each point on each graph.

22 Each graph displays the chromosomal location and least common allele frequency of each  
23 polymorphism of one or more polymorphisms. Each polymorphism displayed on each graph is assigned  
24 a graph location on each graph.

25 The graph location of each polymorphism displayed on each graph is typical of the use of x-y plots. The  
26 graph location assigned to each polymorphism on each graph is a point. The chromosomal location  
27 coordinate of the point assigned as the graph location to any one polymorphism is equal (or  
28 approximately equal) to the chromosomal location of the polymorphism. And the frequency coordinate  
29 of the point assigned as the graph location to any one polymorphism is equal (or approximately equal)  
30 to the least common allele frequency of the polymorphism.

31 The apparatus comprise means for displaying one or more two-dimensional graphs. Each graph  
32 comprises, the identity and graph location of one or more polymorphisms assigned a location on each  
33 graph. And the apparatus comprise means for displaying one or more graphs wherein the viewer  
34 chooses the species, population, polymorphisms, span of the frequency axis and span of the  
35 chromosomal location axis of the one or more graphs ; in versions of the apparatus, the means of this  
36 sentence comprises a computer.

37 The apparatus comprise means for storing and updating data on the chromosomal location and least  
38 common allele frequency of one or more polymorphisms of one or more populations of one or more  
39 species and means for storing chromosomal location and least common allele frequency data on newly  
40 discovered polymorphisms.

Versions of the apparatus comprise means for printing each of the one or more graphs.

## Theory of Operation / Best Mode

### Systematically Varying Both Marker Chromosomal Location and Marker Allele Frequency of Markers in Linkage Studies

The inventor's calculations and observations have demonstrated the increased power of the TDT in more common, less optimal situations when a bi-allelic marker and bi-allelic gene have (1) similar but not identical allele frequencies and (2) the marker and gene are in some degree of linkage disequilibrium. Thus, for a typical linkage study using bi-allelic markers and an association based linkage test, **to increase the likelihood of both criteria (1) and (2) occurring for one or more markers, so as to increase the power of an association based linkage test in a linkage study, the bi-allelic markers used in the study are chosen so that the least common allele frequencies of the markers vary systematically over a range or subrange of least common allele frequency AND the chromosomal location of the markers vary systematically over one or more chromosomes or chromosomal regions. And the bi-allelic markers are chosen so that the markers' chromosomal locations and least common allele frequencies vary systematically in an essentially independent manner.**

(In the Theory of Operation/ Best Mode Section the traditional symbol used in scientific papers for the disequilibrium coefficient,  $\delta$ , is used. This should not be confused with the symbol  $\delta$  used for the covering distance in the remainder of the application. The symbol  $d$  is used for the disequilibrium coefficient in the sections of the application other than the Theory of Operation/Best Mode Section.)

The theory of operation is based on the mathematical observation that the TDT and other association-based tests for linkage are increased in power as the frequencies of the disease-causing allele of a bi-allelic gene and the positively associated allele of a linked bi-allelic marker become similar in magnitude. The inventor made this observation as a result of deriving the equation shown below for  $P_t$  (this is Equation 2 in the unpublished manuscript submitted for publication in December 1996 and in

published paper by RE McGinnis in the Annals of Human Genetics vol 62, pp. 159-179, 1998).

$$P_t = .5 + (1 - 2\theta) \left[ \frac{c_1 c_4 - c_2 c_3}{H} \right] \left\{ p^2 \left( \frac{\alpha^2 - \beta^2}{4} \right) + 2p(1-p) \left( \frac{(\alpha + \beta)^2 - (\beta + \gamma)^2}{16} \right) + (1-p)^2 \left( \frac{\beta^2 - \gamma^2}{4} \right) \right\}$$

Equation 2

$P_t$  may be regarded as the size of the "signal" which is given by the TDT to indicate that a tested marker is linked to a disease-causing gene. The more  $P_t$  is elevated above 0.5 (baseline), the greater is the evidence for linkage or "power" provided by the association-based linkage test known as the TDT.

Table 2 in the unpublished manuscript filed with previous US Provisional Patent Applications (see below) illustrates how signal strength increases substantially as the frequencies of disease-causing allele and positively associated marker allele become similar in magnitude. As noted on pages 24 and 25 of the unpublished manuscript (see below), Table 2 assumes that the frequency ( $p$ )



PCT/US 99/04376

IPEA/US 24 MAY 2000

39

of the disease-causing allele is fixed at  $p=.1$  while the frequency ( $m$ ) of the positively associated marker allele varies ( $m=.5, .3, .2, .1, .05$ ). Note that when the level of disequilibrium (or association) between the bi-allelic marker and bi-allelic disease gene is fixed (in this case either  $\delta=\delta_{\max}$  or  $\delta=\frac{1}{2}\delta_{\max}$ ), the signal strength of  $P_t$  progressively increases as  $m$  decreases from  $m=.5$  to  $m=.1$  (the same frequency as the disease allele, i.e.,  $p=.1$ ). For example, in the section of Table 2 for  $r=.5$ , note that when  $\delta=\frac{1}{2}\delta_{\max}$ ,  $P_t$  is .548 at  $m=.5$  and then steadily increases to .572 ( $m=.3$ ), .597 ( $m=.2$ ), .648 ( $m=.1$ ) and then starts to decrease again as  $m$  departs from  $m=p=.1$  (i.e.  $P_t=.636$  at  $m=.05$ ). As noted on pages 24-25 (below) of the unpublished manuscript, the TDT chi-square statistic (assuming a sample size of 200 families) is such that the signal strength at  $m=.5$  ( $P_t=.548$ ) does not produce a statistically significant evidence for linkage ( $p\text{-value} > 0.05$ ) while the doubling of signal strength at  $m=.2$  ( $P_t=.597$ ) produces very strong statistical evidence for linkage by the TDT ( $p\text{-value} < 0.005$ ). This sort of substantial increase in power is also true of other association-based linkage tests as the frequencies of the disease-causing allele and associated marker allele become more similar in magnitude.

SCANNED #

14

AMENDED SHEET

1 Table 2(Footnotes for Table 2 are on next page)

2 Effect of penetrance ratio (r), disequilibrium ( $\delta$ ) and marker heterozygosity (m) on magnitude  
3 of  $P_t$  and  $P_s$

			Magnitude of $P_t$			Magnitude of $P_s$		
			$\delta_{\max}^a$	$\frac{1}{2} \delta_{\max}^b$	$\delta=0$	$\delta_{\max}^a$	$\frac{1}{2} \delta_{\max}^b$	$\delta=0$
7	r=2	m=.5	.526	.513	.500	.505	.505	.504
8		m=.3	.541	.521	.500	.508	.506	.504
9		m=.2	.558	.531	.500	.511	.508	.504
10		m=.1	.595	.555	.500	.518	.512	.504
11		m=.05	.589	.552	.500	.517	.511	.504
13	r=5	m=.5	.596	.548	.500	.543	.540	.539
14		m=.3	.633	.572	.500	.561	.548	.539
15		m=.2	.666	.597	.500	.575	.556	.539
16		m=.1	.719	.648	.500	.600	.573	.539
17		m=.05	.696	.636	.500	.589	.571	.539
19	r=10	m=.5	.656	.577	.500	.595	.587	.584
20		m=.3	.702	.612	.500	.623	.600	.584
21		m=.2	.736	.644	.500	.644	.612	.584
22		m=.1	.785	.703	.500	.673	.635	.584
23		m=.05	.750	.684	.500	.652	.628	.584
25	r= $\infty$	m=.5	.740	.617	.500	.700	.680	.673
26		m=.3	.791	.663	.500	.743	.700	.673
27		m=.2	.826	.703	.500	.772	.716	.673
28		m=.1	.870	.770	.500	.807	.744	.673
29		m=.05	.816	.741	.500	.763	.730	.673

1 Footnotes for Table 2

2 a.<sup>b</sup>Value of  $\delta$  that is maximal ( $\delta_{\max}$ ) and half-maximal ( $\frac{1}{2}\delta_{\max}$ ), as determined by the  
3 heterozygosity of the marker ( $m$ ) and disease locus ( $p=.1$ )  
4

5 Importance of disequilibrium and marker heterozygosity (i.e. marker allele frequency) in  
6 detecting linkage

7 When the heterozygosity (i.e. allele frequencies) of a bi-allelic marker and bi-allelic  
8 disease locus are fixed, ( $P_S = .5$ ) and  $|P_t - .5|$  are both maximized at the most positive or most  
9 negative possible value of  $\delta$  ( $\delta_{\max}$ ,  $\delta_{\min}$ ), as demonstrated in the published paper. This  
10 maximization of  $\chi^2_{\text{asp}}$  and  $\chi^2_{\text{tdt}}$  is intimately connected to  $M_S$  and  $M_t$  (defined in equations 1  
11 and 2) since: (a) these are the only two factors in  $P_S$  and  $P_t$  that are influenced by  $\delta$  and (b)  $M_S$   
12 and  $|M_t|$  are maximal and equal to each other when  $\delta$  is extreme ( $\delta_{\max}$  or  $\delta_{\min}$ ). Furthermore, as  
13 explained in the published paper,  $M_S$  is a measure of the proportion of informative (A/B)  
14 parents who are also informative (D/d) at the disease locus. Therefore, maximizing  $M_S$  (and,  
15 by implication,  $|M_t|$ ) is equivalent to *minimizing* the proportion of A/B parents who are  
16 homozygous (D/D or d/d) at the disease locus. Such homozygous D/D or d/d parents  
17 contribute evidence *against* linkage since they transmit marker alleles A and B to affected  
18 offspring with equal probability; thus, minimizing their proportion among A/B parents being  
19 tested for linkage has the effect of maximizing  $\chi^2_{\text{asp}}$  and  $\chi^2_{\text{tdt}}$ .

20 Nevertheless, when bi-allelic markers have a specific (i.e. fixed) heterozygosity  
21 different from that of a bi-allelic disease locus, some A/B parents must be homozygous at the  
22 disease locus, even when  $\delta$  is extreme. But if marker heterozygosity is variable, the proportion  
23 of A/B parents who are D/D or d/d approaches *zero* as marker heterozygosity approaches that  
24 of the disease locus and as  $\delta$  approaches  $\delta_{\max}$  or  $\delta_{\min}$ . Consequently, the most extreme values  
25 of  $P_t$  and  $P_S$ , and highest values of  $\chi^2_{\text{tdt}}$  and  $\chi^2_{\text{asp}}$  are found when marker and disease locus  
26 have equal heterozygosity and  $\delta = \delta_{\max}$  or  $\delta = \delta_{\min}$ .

27

28 Example illustrating the importance of marker heterozygosity (i.e. allele frequency)

To illustrate the importance of marker heterozygosity and disequilibrium, Table 2 shows  $P_t$  and  $P_s$  values when the frequency ( $p$ ) of disease allele  $D$  is constant at 0.1, but the frequency ( $m$ ) of marker allele  $A$  varies between  $m=.5$  (maximum marker heterozygosity) and  $m=.1$  (equal heterozygosity at marker and disease loci). The table assumes mode of inheritance is additive, and separate sections of the table show the results when the penetrance ratio ( $r$ ) is 2, 5, 10 or  $\infty$ . For each value of  $r$ , an individual line in the table represents constant marker heterozygosity ( $m=.5, .3, .2$ , or  $.1$ ) and from left-to-right on each line, one sees  $P_t$  and  $P_s$  values when  $\delta=\delta_{\max}$ ,  $\delta=\frac{1}{2}\delta_{\max}$ , and  $\delta=0$ , the value of  $\delta_{\max}$  being determined by the particular values of  $m$  and  $p$  [ $\delta_{\max}=p(1-m)$ ]. As noted in Appendix I of the published paper, when  $p<m$  and  $p<(1-m)$ , as in this example, the most extreme values of  $P_t$  and  $P_s$  must occur at  $\delta=\delta_{\max}$ . This can be seen in each line of the table by the steady increase in both  $P_t$  and  $P_s$  as one moves from  $\delta=0$  to  $\delta=\delta_{\max}$ , with every line also showing  $P_t > P_s$  at  $\delta=\delta_{\max}$  and most lines showing  $P_t > P_s$  at  $\delta=\frac{1}{2}\delta_{\max}$ .

Most remarkable, however, are the sizeable increases in  $P_s$  and even greater increases in  $P_t$  as marker heterozygosity drops toward the heterozygosity of the disease locus ( $m \rightarrow .1$ ). A typical example is at  $r=5$  and  $\delta=\frac{1}{2}\delta_{\max}$  where the table shows  $P_t=.548$  at maximum marker heterozygosity ( $m=.5$ ) and  $P_t=.597$  or  $.648$  for  $m=.2$  or  $.1$ , respectively. The impact of such an increase in  $P_t$  can be understood by calculating  $\chi^2_{\text{tdt}}$  for  $P_t=.548$  ( $m=.5$ ) and for  $P_t=.597$  ( $m=.2$ ) assuming a data set of 200 families each with two affected sibs. Based on the expression for  $\frac{H}{F}$ , I calculate the proportion of A/B parents to be .50 and .39 when  $m=.5$  and  $.2$ , respectively. So for  $m=.5$ , there would be  $.5 \times 400 \times 2 = 400$  informative transmissions to affected offspring with transmissions of allele A totaling  $.548 \times 400 = 219$ , thus implying  $\chi^2_{\text{tdt}} = \frac{38^2}{400} = 3.61$ ,  $p<0.1$ . For  $m=.2$ , there would be  $.39 \times 400 \times 2 = 312$  informative transmissions of which  $.597 \times 312 = 186$  would be transmissions of allele A yielding  $\chi^2_{\text{tdt}} = \frac{60^2}{312} = 11.54$ ,  $p<0.005$ .

This example is typical, and highlights perhaps the most important finding of this paper; namely the importance of using bi-allelic markers with heterozygosity similar to that of a bi-allelic disease locus. Indeed, since a majority of susceptibility loci may be bi-allelic, the

judicious use of bi-allelic markers of both high, medium, and low heterozygosity may be crucial in order to initially detect and replicate linkages to loci conferring modest disease risk.

**Best Mode:**

Method for locating disease causing polymorphism using biallelic linkage analysis

Objective :To test, by association-based linkage analysis (e.g., by TDT), whether a disease-causing polymorphism is located on a particular chromosome (e.g., human chromosome 4) or within a particular subregion of that chromosome.

**PART 1 - Steps in conducting the association-based linkage test**

**Step 1**

To conduct the test, first divide the chromosome or subregion of interest into segments that are short enough that polymorphisms within each segment are likely to be in linkage disequilibrium with each other. The division of a chromosome or subregion of interest into "segments" is conceptual (*not* physical) and is based on chromosomal maps such as those provided by the Whitehead Institute or Marshfield Foundation for Biomedical Research. Although disequilibrium has been observed in Finnish populations between polymorphisms that are 7 to 10 centimorgans (cM) apart, the chromosomal segments for searching for disease-causing polymorphisms in more genetically heterogeneous populations should be less than 1 cM long (e.g., 250,000 base pairs long). These chromosomal segments might or might not overlap each other (i.e., share some of their length in common); but the set of chromosomal segments should completely cover the entire chromosome or entire subregion of interest, so that a disease-causing polymorphism located anywhere on the chromosome or anywhere in the subregion of interest will be detected by the test.

**Step 2**

It is well known that increased disequilibrium between a marker and linked disease locus increases evidence for linkage provided by association-based linkage tests such as the TDT. However, what has not been recognized is that the specific allele frequencies of the marker locus can also have an enormous impact on the strength of evidence for linkage. I

showed this by analyzing equation 2 for  $P_t$ . Specifically, when a bi-allelic marker is in linkage disequilibrium with a bi-allelic disease locus, the strength of evidence for linkage provided by the TDT is *greatly* increased if the bi-allelic marker and bi-allelic disease locus have similar allele frequencies.

This phenomenon is illustrated by Table 2 and explained above. For example, suppose as noted above, that the susceptibility allele ("allele D") of a bi-allelic disease locus has a frequency of 0.1 and further suppose that the disease locus is in half-maximal positive disequilibrium with a bi-allelic marker ( $\delta = \frac{1}{2} \delta_{\max}$ ). As noted above,  $\chi^2_{\text{TDT}}$  will equal only 3.61 ( $p < 0.1$ ) if the frequency of the less common marker allele is 0.5; but if the frequency of the less common marker allele is 0.2 (and hence much closer to the frequency of allele D) then  $\chi^2_{\text{TDT}}$  will equal 11.54, thus providing much stronger evidence for linkage ( $p < 0.005$ ).

Therefore, in searching for association-based linkage to a bi-allelic disease locus within each of the aforementioned chromosomal segments (see step 1), it is crucial to identify and test (e.g., by TDT) bi-allelic markers within each segment that have a broad range of allele frequencies. An unidentified bi-allelic disease locus could have allele frequencies close to 0.5/0.5, 0.4/0.6, 0.3/0.7, 0.2/0.8, 0.1/0.9 or below 0.1/above 0.9; hence, it is crucial to test bi-allelic markers with frequencies near 0.5/0.5 and near 0.1/0.9 as well as test others with allele frequencies that fall at regular increments between the extremes of 0.5/0.5 and 0.1/0.9. By testing bi-allelic markers with a broad range of allele frequencies that are spaced at regular intervals between 0.5/0.5 and 0.1/0.9, one is assured of testing some bi-allelic markers whose two allele frequencies are reasonably close to the allele frequencies of an unknown bi-allelic disease locus.

Thus, for step 2, within each chromosomal segment, subsets of bi-allelic markers should be identified. Each subset contains only bi-allelic markers having approximately the same allele frequencies. For example, subset A contains only markers whose less common allele has a population frequency of about 0.1. Similarly, subsets B, C, D, and E contain only bi-allelic markers whose less common allele has a frequency of approximately 0.2, 0.3, 0.4, and 0.5, respectively. In other versions of the invention the number of subsets is greater or less than five, and the approximate allele frequency of the less common bi-allele of subsets is other than about 0.1, 0.2, 0.3, 0.4 or 0.5 and is expected to be more than one decimal long since allele frequencies from real populations are rarely round numbers. However, the crucial point is that each subset should contain only bi-allelic markers belonging to one chromosomal segment and the frequency of the less common allele of each subset member should be

approximately the same (i.e., the *difference* between the frequencies of the less common allele of any two subset members should not exceed 0.15). Also crucial, as I emphasized above, is that the *group* of subsets for each chromosomal segment represent frequencies near the extremes of 0.5/0.5 and 0.1/0.9 as well as represent bi-allele frequencies between these two extremes that are approximately evenly spaced as *illustrated* by the group of subsets referred to above as A, B, C, D and E.

### Step 3

In step 2, I described the importance of testing subsets of bi-allelic markers having approximately the same frequencies for their two alleles. Here I further delineate the characteristics of the markers that should be chosen for each subset by noting why it is important that each subset contain more than one bi-allelic marker. Even though a particular bi-allelic marker has allele frequencies that are similar to those of a closely linked bi-allelic disease locus, the marker may not be in strong positive disequilibrium with the disease locus. If disequilibrium is minimal, the marker will not show strong evidence for linkage under the TDT or any other association-based linkage test, *even though the bi-allelic marker and disease locus have similar allele frequencies*.

Hence, it is important that each subset contain multiple bi-allelic markers so that there is increased likelihood that at least one of the markers will be in reasonably strong disequilibrium with a closely linked bi-allelic disease locus. Beyond the cardinal criterion that all bi-allelic markers in a subset have similar allele frequencies, an additional criteria for selecting markers to belong to a subset is that the chosen bi-allelic markers *should not be in extreme positive disequilibrium with each other*.

The reason for this is as follows: According to standard usage, the disequilibrium coefficient ( $\delta$ ) is defined by the equation  $\delta = f(AB) - f(A)f(B)$  where  $f(A)$  and  $f(B)$  may be defined as the frequencies of the less common allele (denoted A and B) of two bi-allelic loci belonging to the same subset and  $f(AB)$  is the population frequency of the AB haplotype. Since the two markers belong to the same subset, we may assume that  $f(A)=f(B)=q$ ; hence the maximum positive value of  $\delta$  (denoted  $\delta_{\max}$ ) is  $\delta=q-q^2$ . This maximum positive  $\delta$  value (i.e. maximum "positive disequilibrium") occurs if every chromosome that carries allele A also carries allele B, and if every chromosome that carries allele not-A also carries allele not-B. Hence, when two bi-allelic markers with similar allele frequencies are in extreme positive disequilibrium with each other (i.e.,  $\delta$  is approximately equal to  $\delta_{\max}$ ), the two loci provide

the nearly identical information with respect to their linkage and association with a third polymorphism such as a disease locus. Hence one of the two bi-allelic markers would provide no additional information and its inclusion in the subset would not increase the likelihood of detecting linkage and association to a nearby disease locus.

Therefore, bi-allelic markers belonging to the same chromosomal segment and subset should not only have similar allele frequencies, the  $\delta$  value between *each pair* of bi-allelic markers in the same subset should be substantially less than  $\delta_{\max} = q \cdot q^2$ . This assures that every bi-allelic polymorphism belonging to the subset provides much new (i.e. non-redundant) information about linkage and association to any nearby bi-allelic disease locus: thus testing each bi-allelic marker in the subset would increase the likelihood of detecting linkage to a disease locus.

#### Step4: Test for linkage

To test for (association-based) linkage to a bi-allelic disease locus, each bi-allelic marker in each subset from each chromosomal segment is tested *individually* by using the TDT, AFBAC method or other family-based linkage test. To conduct these tests for a particular marker, members of nuclear families (most especially parents, and any children who manifest disease) are genotyped at the marker being tested and the genotypes are then evaluated according to the TDT, AFBAC method or other family-based linkage/association test (for description of TDT and AFBAC, see Spielman et al, Am J of Human Genetics 52:506-516 (1993) and Thomson, Am J Human Genetics 57:487-498 (1995)). Alternatively, linkage and association is tested for each marker in each subset from each segment by genotyping individuals with disease and related or unrelated normal controls at each marker to be tested. (End of best mode example)

#### Further Information

(Step 3 is not essential for the operation or utility of this version of the invention. In this best mode example, the least common allele frequency subrange 0.1 to 0.5 is used. In versions of the invention similar to the best mode, versions of the invention are operable and have utility for any subrange of the least common allele frequency range 0 to 0.5. In addition, rather than genotyping DNA from single individuals in step 4, in some versions of the invention each marker in each subset from each segment is tested for association with disease by evaluating DNA from pooled samples.)



## 1 PART 2 - Physical implementation of the above test

2  
3 Silicon chips or glass slides with arrays of oligonucleotides for testing bi-allelic markers

4 Companies like Affymetrix<sup>™</sup> are using silicon chips or glass slides to genotype DNA  
5 from one individual at thousands of bi-allelic markers. Each silicon chip or glass slide is  
6 divided into a grid or 2-dimensional matrix that contains thousands of cells. To the surface of  
7 each cell is attached multiple copies of a unique oligonucleotide whose sequence is  
8 complementary (type (1)) to one of the two alleles of a particular bi-allelic marker. Thus, DNA  
9 from an individual who carries the allele hybridizes to the cell with substantially greater  
10 affinity than does the alternate bi-allele. The degree of hybridization generates a signal and  
11 enables the genotype of the individual to be inferred for that particular bi-allelic polymorphism  
12 [i.e., the individual is homozygous (++) , heterozygous (+-), or homozygous (--)]. In some  
13 applications, it is crucial to attach oligonucleotides corresponding to each allele of a bi-allelic  
14 polymorphism in adjacent cells so that the relative (i.e. local) intensity of hybridization in the  
15 adjacent cells can be compared, thus facilitating inference of the individual's correct genotype  
16 (++, +-, or --).

17 In using this silicon chip or glass slide technology to test for linkage and association,  
18 the ideas detailed in PART 1 indicate how the oligonucleotides that are attached to the cells of  
19 the silicon chip or glass slide should be chosen. To scan a particular chromosome or  
20 chromosomal region for a bi-allelic disease locus, the chromosome or chromosomal region  
21 should be subdivided into segments as described in Step 1 above. For each segment, subsets of  
22 bi-allelic markers having the properties detailed in PART 1 above should be identified. The  
23 DNA of select individuals (see "Test for linkage" - above) should then be assayed at each bi-  
24 allelic marker in every chromosomal segment and in every subset of markers belonging to the  
25 segment. This would be accomplished by attaching an oligonucleotide corresponding to one of  
26 the marker's two alleles to a particular (i.e. known) cell on the silicon chip or slide. To  
27 enhance assignment of accurate genotypes, it may also be advisable to attach an  
28 oligonucleotide corresponding to the second allele of the bi-allelic marker in an adjacent cell as  
29 mentioned in the previous paragraph.

30 Industrial Applicability

31 Versions of the present invention are useful for locating trait causing genes and polymorphisms such as  
32 human disease genes and polymorphisms. Versions of the invention could be used to find the cure for  
33 human disease. The making and use of versions of the invention should be clear to a person of skill in  
34 the art after reading the description.

35 Scope of the Invention

While the description contains many specificities, these should not be construed as limitations on the scope of the invention, but rather as exemplifications of versions of the invention.

Accordingly the scope of the invention should be determined not by the specific versions described alone, but also by the claims and their legal equivalents and also by any future claims drawn to the invention and future descriptions of versions of the invention.

#### Notes:

The reader's attention is directed to the following papers which are open to the public and are herein incorporated by reference: (1) McGinnis, Ewens & Spielman, Genetic Epidemiology 1995 ; 12(6) 637-40. (2) RE McGinnis Annals of Human Genetics vol 62, pp. 159-179, 1998. The papers in the endnotes below are incorporated herein by reference.

<sup>i</sup> Weighing DNA for Fast Genetic Diagnosis, Science, March 27, 1998, vol. 279, pp. 2044-2045.

<sup>ii</sup> Spielman, R.S. and Ewens, W.J. The TDT and Other Family-Based Tests for Linkage Disequilibrium and Association, American Journal of Human Genetics, 59: 983-989, 1996.

<sup>iii</sup> "Mathematical Theory of Optimization" The New Encyclopedia Britannica, 15<sup>th</sup> edition, vol. 25, pp. 217-221.

<sup>iv</sup> American Journal of Human Genetics, vol. 57: 439-454, 1995.

<sup>v</sup> American Journal of Human Genetics, vol. 58: 1347-1363, 1996.

<sup>vi</sup> Human Heredity, vol. 44, pp. 225-237, 1994.

<sup>vii</sup> Human Heredity, vol. 46, pp. 226-235, 1996.

<sup>viii</sup> Accessing Genetic Information with High-Density DNA Arrays, Mark Chee, et al. Science, vol 274, Oct. 25, 1996, pp. 610 - 614.

<sup>ix</sup> Large Scale Identification, Mapping, and Genotyping of Single-Nucleotide Polymorphisms in the Human Genome, Wang, et. al., Science, May 15, 1998, vol 280, pp. 1077-1081.

<sup>x</sup> (1) Schuster, H. et al (1995) Nature Genetics, 13(1) : 98 - 100.

(2) Gyapay, G. et al (1994) Nature Genetics, 7: 246-339.

<sup>xi</sup> Some versions of oligonucleotide technology and its uses to obtain genetic information are included in the following papers:

(1) Accessing Genetic Information with High-Density DNA Arrays, Mark Chee, et al. Science, vol 274, Oct. 25, 1996, pp. 610 - 614.

(2) Genetic analysis of amplified DNA with immobilized sequence-specific oligonucleotide probes, Saiki, et al. Proc Natl Acad Sci USA vol 86, pp. 6230-6234.

(3) Allele-specific enzymatic amplification of  $\beta$ -globin genomic DNA for diagnosis of sickle cell anemia, Wu, et al., Proc Natl Acad Sci USA vol 86 pp 2757-2760.

(4) Automated DNA diagnostics using an Elisa-based oligonucleotide ligation assay, Nickerson, et al., Proc Natl Acad Sci USA vol 87, pp. 8923-8927.

(5) Genetic analysis of amplified DNA with immobilized sequence specific oligonucleotide probes, Saiki, et al., Proc Natl Acad Sci USA vol 86 pp 6230 - 6234.

<sup>xii</sup> Padlock Probes: Circularizing Oligonucleotides for Localized DNA Detection, Science, Sept. 30, 1994, vol. 265, pp. 2085-2088.

<sup>xiii</sup> SNP attack on complex traits, Nature Genetics, Nov. 1998, vol. 20 no. 3, pp. 217-218.

PCT/US 99/04376

IPEA/US 22 MAY 2000

PCT/US99/04376

49  
95  
Claims

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34

What is claimed

1. An invention substantially as described in the description
2. An invention substantially as described and illustrated in the description.
3. A process for identifying one or more bi-allelic markers linked to a bi-allelic genetic characteristic gene in a species of creatures, comprising the steps of :
  - a)choosing two or more bi-allelic covering markers so that a CL-F region is systematically covered by the two or more covering markers, the CL-F region being a collection of points on a two-dimensional plane, the two-dimensional plane having the two orthogonal dimensions of chromosomal location and least common allele frequency,
  - b)choosing a statistical linkage test based on allelic association for each covering marker;
  - c)choosing a sample of individuals for each covering marker ;
  - d)obtaining genotype data/sample allele frequency data for each covering marker and the sample chosen for each covering marker, and obtaining phenotype status data for the genetic characteristic for each individual in the sample chosen for each covering marker;
  - e)calculating evidence for linkage between each covering marker and the gene using the statistical linkage test based on allelic association chosen for each covering marker and the genotype data/sample allele frequency data for each covering marker and using the phenotype status data for the genetic characteristic for each individual in the sample chosen for each covering marker obtained in d); and
  - f)identifying those covering markers as linked to the genetic characteristic gene which show evidence for linkage based on the calculations of step e)
4. A process as in claim 3, wherein the CL-F region is N covered to within a CL-F distance  $\delta$  by the two or more bi-allelic covering markers, so that each point in the region is within the CL-F distance  $\delta$  of N or more of the covering markers, wherein  $\delta$  is equal to about  $[\delta_{CL}, 0.25]$  or the equivalent thereof,  $\delta_{CL}$  is equal to the largest chromosomal length, computed by any method, for which linkage disequilibrium has

AMENDED SHEET

50

~~76~~

1 been observed between any polymorphisms in any population of the species, N is an integer greater  
2 than or equal to 1.

3 5. A process as in claim 4, wherein the CL-F region includes 81 percent or more of the centerpoints of  
4 the matrix centerpoint lattice of a CL-F matrix, the number of cells in the matrix being greater than or  
5 equal to three, wherein the matrix has R rows and C columns, each cell of the matrix being of length  $L_{MC}$   
6 and width  $W_{MC}$ , and  $L_{MC}$  being less than or equal to about  $2\delta_{CL}$ , and  $W_{MC}$  being less than or equal to 0.5,  
7  $\delta_{CL}$  is equal to the largest chromosomal length, computed by any method, for which linkage  
8 disequilibrium has been observed between any polymorphisms in any population of the species, there  
9 being N or more covering markers in each cell of 81 percent or more of the cells of the matrix, N is an  
10 integer greater than or equal to 1; the covering markers being distributed over a chromosomal region of  
11 interest, the region of interest being approximately the smallest chromosome interval that contains all of  
12 the covering markers, and the covering markers comprising essentially less than all of the  
13 polymorphisms in the region of interest

14 6. A process as in claim 5, wherein the covering markers are substantially nonevenly distributed across  
15 a chromosome or a chromosomal segment.

16 7. A process as in claim 5, wherein the covering markers are substantially evenly distributed across a  
17 chromosome or a chromosomal segment, and wherein there is a subgroup of one or more of the  
18 covering markers, and each of the markers in the subgroup is chosen without substantial preference for  
19 the least common allele frequency of each of the markers in the subgroup being close to 0.5, and the  
20 number of covering markers in the subgroup is about 5 percent or more of the total number of covering  
21 markers.

22 8. A process as in claim 5, wherein the covering markers are substantially evenly distributed across a  
23 chromosome or a chromosomal segment, and wherein there is a subgroup of one or more of the  
24 covering markers, and each of the markers in the subgroup is chosen without substantial preference for  
25 the least common allele frequency of each of the markers in the subgroup being close to 0.5.

26 9. A process as in claim 5, wherein the covering markers are substantially evenly distributed across a  
27 chromosome or a chromosomal segment, wherein (1) the average chromosomal intermarker distance of  
28 the covering markers is greater than 2 cM or the equivalent thereof and the least common allele  
29 frequency of one or more of the covering markers is less than 0.3, or wherein (2) the least common  
30 allele frequency of one or more of the covering markers is less than 0.2, or wherein (3) the average  
31 chromosomal intermarker distance of the covering markers is greater than 10 cM or the equivalent  
32 thereof

33 10. A process as in claim 5, wherein the covering markers are substantially evenly distributed across a  
34 chromosome or a chromosomal segment, wherein the average chromosomal intermarker distance of  
35 the covering markers is less than or equal to 2 cM or the equivalent thereof, and wherein the conditional

1 probability the covering markers were chosen essentially randomly from substantially the known set of  
2 bi-allelic markers with least common allele frequencies between 0.2 inclusive and 0.5 inclusive is less  
3 than or equal to 10 percent; wherein the conditional probability is substantially conditional on (1) the  
4 approximate chromosomal distribution of the covering markers, (2) the marker type of each covering  
5 marker and (3) there being N or more covering markers in each cell of 81 percent or more of the cells of  
6 the matrix.

7 11 A process as in claim 5, wherein the covering markers are substantially evenly distributed across a  
8 chromosome or a chromosomal segment, wherein the average chromosomal intermarker distance of  
9 the covering markers is greater than 2 cM or the equivalent thereof; and wherein the conditional  
10 probability the covering markers were chosen essentially randomly from substantially the known set of  
11 bi-allelic markers with least common allele frequencies between 0.3 inclusive and 0.5 inclusive is less  
12 than or equal to 10 percent; wherein the conditional probability is substantially conditional on (1) the  
13 approximate chromosomal distribution of the covering markers, (2) the marker type of each covering  
14 marker and (3) there being N or more covering markers in each cell of 81 percent or more of the cells of  
15 the matrix

16 12. A process as in claim 10, wherein the chromosome or the chromosomal segment consists  
17 essentially of a set of nonoverlapping chromosome segments of substantially equal length, and wherein  
18 each chromosome segment of the set has two or less covering markers located thereon, wherein one  
19 and only one covering marker is located on each of 80 percent or more of the chromosome segments of  
20 the set, and wherein zero or two and only two covering markers are located on each of 20 percent or  
21 less of the chromosome segments of the set, and wherein each chromosome segment that borders a  
22 chromosome segment with zero covering markers located thereon has only one or two covering  
23 markers located thereon, and wherein each chromosome segment that borders a chromosome segment  
24 with two covering markers located thereon has only one or zero covering markers located thereon; and  
25 wherein the conditional probability the covering markers were chosen essentially randomly from  
26 substantially the known set of bi-allelic markers with least common allele frequencies between 0.2  
27 inclusive and 0.5 inclusive is less than or equal to 10 percent; wherein the conditional probability is  
28 substantially conditional on (1) the chromosomal distribution of the covering markers on the  
29 chromosome segments of the set, (2) the marker type of each covering marker and (3) there being N or  
30 more covering markers in each cell of 81 percent or more of the cells of the matrix

31 13 A process as in claim 11, wherein the chromosome or the chromosomal segment consists  
32 essentially of a set of nonoverlapping chromosome segments of substantially equal length, and wherein  
33 each chromosome segment of the set has two or less covering markers located thereon, wherein one  
34 and only one covering marker is located on each of 80 percent or more of the chromosome segments of  
35 the set, and wherein zero or two and only two covering markers are located on each of 20 percent or

52  
~~78~~

1 less of the chromosome segments of the set, and wherein each chromosome segment that borders a  
2 chromosome segment with zero covering markers located thereon has only one or two covering  
3 markers located thereon, and wherein each chromosome segment that borders a chromosome segment  
4 with two covering markers located thereon has only one or zero covering markers located thereon; and  
5 wherein the conditional probability the covering markers were chosen essentially randomly from  
6 substantially the known set of bi-allelic markers with least common allele frequencies between 0.3  
7 inclusive and 0.5 inclusive is less than or equal to 10 percent, wherein the conditional probability is  
8 substantially conditional on (1) the chromosomal distribution of the covering markers on the  
9 chromosome segments of the set, (2) the marker type of each covering marker and (3) there being N or  
10 more covering markers in each cell of 81 percent or more of the cells of the matrix

11 14 A process as in claim 5, wherein the covering markers are substantially evenly distributed across a  
12 chromosome or a chromosomal segment, wherein the average chromosomal intermarker distance of  
13 the covering markers is less than or equal to 2 cM or the equivalent thereof; and wherein collection C is  
14 essentially the collection of known groups of bi-allelic markers with least common allele frequencies  
15 between 0.2 inclusive and 0.5 inclusive, each group in the collection being substantially similar to the  
16 covering markers as a group; wherein a group of bi-allelic markers is a member of collection C if and  
17 only if the group substantially meets criteria (1), (2), (3) and (4); wherein criterion (1) is, each marker in  
18 the group is chosen from substantially the known set of bi-allelic markers with least common allele  
19 frequencies between 0.2 inclusive and 0.5 inclusive, wherein criterion (2) is, the number of markers in  
20 the group is the same as the number of covering markers, wherein criterion (3) is, the chromosomal  
21 distribution of the group of markers and the covering markers is substantially similar, and wherein  
22 criterion (4) is, the marker type of each group marker and the covering marker with substantially the  
23 same chromosomal location is the same; wherein a group that is a member of collection C meets  
24 criterion (5) if and only if (5) there are N or more of the group markers in each cell of 81 percent or more  
25 of the cells of a CL-F matrix with cells of length  $L_{MC}$  and width  $W_{MC}$  in R rows and C columns; wherein P  
26 is essentially the proportion of groups in collection C that meet criterion (5); wherein P is less than 90  
27 percent.

28 15. A process as in claim 5, wherein the covering markers are substantially evenly distributed across a  
29 chromosome or a chromosomal segment, wherein the average chromosomal intermarker distance of  
30 the covering markers is greater than 2 cM or the equivalent thereof, and wherein collection C is  
31 essentially the collection of known groups of bi-allelic markers with least common allele frequencies  
32 between 0.3 inclusive and 0.5 inclusive, each group in the collection being substantially similar to the  
33 covering markers as a group, wherein a group of bi-allelic markers is a member of collection C if and  
34 only if the group substantially meets criteria (1), (2), (3) and (4); wherein criterion (1) is, each marker in  
35 the group is chosen from substantially the known set of bi-allelic markers with least common allele

53

~~79~~

1 frequencies between 0.3 inclusive and 0.5 inclusive, wherein criterion (2) is, the number of markers in  
2 the group is the same as the number of covering markers, wherein criterion (3) is, the chromosomal  
3 distribution of the group of markers and the covering markers is substantially similar, and wherein  
4 criterion (4) is, the marker type of each group marker and the covering marker with substantially the  
5 same chromosomal location is the same, wherein a group that is a member of collection C meets  
6 criterion (5) if and only if (5) there are N or more of the group markers in each cell of 81 percent or more  
7 of the cells of a CL-F matrix with cells of length  $L_{MC}$  and width  $W_{MC}$  in R rows and C columns, wherein P  
8 is essentially the proportion of groups in collection C that meet criterion (5), wherein P is less than 90  
9 percent

10 16. A process as in claim 5, wherein the covering markers are substantially evenly distributed across a  
11 chromosome or a chromosomal segment, wherein the average chromosomal intermarker distance of  
12 the covering markers is less than or equal to 2 cM or the equivalent thereof, wherein the chromosome or  
13 the chromosomal segment consists essentially of a set of nonoverlapping chromosome segments of  
14 substantially equal length, and wherein each chromosome segment of the set has two or less covering  
15 markers located thereon, wherein one and only one covering marker is located on each of 80 percent or  
16 more of the chromosome segments of the set, and wherein zero or two and only two covering markers  
17 are located on each of 20 percent or less of the chromosome segments of the set, and wherein each  
18 chromosome segment that borders a chromosome segment with zero covering markers located thereon  
19 has only one or two covering markers located thereon, and wherein each chromosome segment that  
20 borders a chromosome segment with two covering markers located thereon has only one or zero  
21 covering markers located thereon; wherein collection D is essentially the collection of known groups of  
22 bi-allelic markers with least common allele frequencies between 0.2 inclusive and 0.5 inclusive, each  
23 group in the collection being substantially similar to the covering markers as a group; wherein a group of  
24 bi-allelic markers is a member of collection D if and only if the group substantially meets criteria (1), (2),  
25 (3) and (4); wherein criterion (1) is, each marker in the group is chosen from substantially the known set  
26 of bi-allelic markers with least common allele frequencies between 0.2 inclusive and 0.5 inclusive,  
27 wherein criterion (2) is, the number of markers in the group is the same as the number of covering  
28 markers, wherein criterion (3) is, the chromosomal distribution of the group of markers and the covering  
29 markers is substantially similar, and wherein criterion (4) is, the marker type of each group marker and  
30 the covering marker with substantially the same chromosomal location is the same, wherein a group  
31 that is a member of collection D meets criterion (5) if and only if (5) there are N or more of the group  
32 markers in each cell of 81 percent or more of the cells of a CL-F matrix with cells of length  $L_{MC}$  and  
33 width  $W_{MC}$  in R rows and C columns; wherein P is essentially the proportion of groups in collection D  
34 that meet criterion (5), wherein P is less than 90 percent.

**AMENDED SHEET**

54  
80

17. A process as in claim 5, wherein the covering markers are substantially evenly distributed across a chromosome or a chromosomal segment, wherein the average chromosomal intermarker distance of the covering markers is greater than 2 cM or the equivalent thereof; wherein the chromosome or the chromosomal segment consists essentially of a set of nonoverlapping chromosome segments of substantially equal length, and wherein each chromosome segment of the set has two or less covering markers located thereon, wherein one and only one covering marker is located on each of 80 percent or more of the chromosome segments of the set, and wherein zero or two and only two covering markers are located on each of 20 percent or less of the chromosome segments of the set, and wherein each chromosome segment that borders a chromosome segment with zero covering markers located thereon has only one or two covering markers located thereon, and wherein each chromosome segment that borders a chromosome segment with two covering markers located thereon has only one or zero covering markers located thereon; wherein collection D is essentially the collection of known groups of bi-allelic markers with least common allele frequencies between 0.3 inclusive and 0.5 inclusive, each group in the collection being substantially similar to the covering markers as a group; wherein a group of bi-allelic markers is a member of collection D if and only if the group substantially meets criteria (1), (2), (3) and (4); wherein criterion (1) is, each marker in the group is chosen from substantially the known set of bi-allelic markers with least common allele frequencies between 0.3 inclusive and 0.5 inclusive, wherein criterion (2) is, the number of markers in the group is the same as the number of covering markers, wherein criterion (3) is, the chromosomal distribution of the group of markers and the covering markers is substantially similar, and wherein criterion (4) is, the marker type of each group marker and the covering marker with substantially the same chromosomal location is the same; wherein a group that is a member of collection D meets criterion (5) if and only if (5) there are N or more of the group markers in each cell of 81 percent or more of the cells of a CL-F matrix with cells having length  $L_{Mc}$  and width  $W_{Mc}$  in R rows and C columns, wherein P is essentially the proportion of groups in collection D that meet criterion (5); wherein P is less than 90 percent.

18. A process as in claim 4, wherein the covering markers meet criterion a) and one or more of the criteria b), c), d), e), f), g), h) or i), wherein

criterion a) is the covering markers are distributed over a chromosomal region of interest, the region of interest being approximately the smallest chromosome interval that contains all of the covering markers, and the covering markers comprise essentially less than all of the polymorphisms in the region of interest,

criterion b) is the covering markers are substantially nonevenly distributed across a chromosome or a chromosomal segment,

**AMENDED SHEET**



1  
2 criterion c) is the covering markers are substantially evenly distributed across a chromosome or a  
3 chromosomal segment, and there is a subgroup of one or more of the covering markers, and each of  
4 the markers in the subgroup is chosen without substantial preference for the least common allele  
5 frequency of each of the markers in the subgroup being close to 0.5, and the number of covering  
6 markers in the subgroup is about 5 percent or more of the total number of covering markers,

7  
8 criterion d) is the covering markers are substantially evenly distributed across a chromosome or a  
9 chromosomal segment, and there is a subgroup of one or more of the covering markers, and each of  
10 the markers in the subgroup is chosen without substantial preference for the least common allele  
11 frequency of each of the markers in the subgroup being close to 0.5,

12  
13 criterion e) is the covering markers are substantially evenly distributed across a chromosome or a  
14 chromosomal segment, wherein (1) the average chromosomal intermarker distance of the covering  
15 markers is greater than 2 cM or the equivalent thereof and the least common allele frequency of one or  
16 more of the covering markers is less than 0.3, or wherein (2) the least common allele frequency of one  
17 or more of the covering markers is less than 0.2, or wherein (3) the average chromosomal intermarker  
18 distance of the covering markers is greater than 10 cM or the equivalent thereof;

19  
20 criterion f) is the covering markers are substantially evenly distributed across a chromosome or a  
21 chromosomal segment, wherein the average chromosomal intermarker distance of the covering markers  
22 is less than or equal to 2 cM or the equivalent thereof, and wherein the conditional probability the  
23 covering markers were chosen essentially randomly from substantially the known set of bi-allelic  
24 markers with least common allele frequencies between 0.2 inclusive and 0.5 inclusive is less than or  
25 equal to 10 percent; wherein the conditional probability is substantially conditional on (1) the  
26 approximate chromosomal distribution of the covering markers, (2) the marker type of each covering  
27 marker and (3) the CL-F region being N covered to within the CL-F distance  $\delta$  by the two or more bi-  
28 allelic covering markers;

29  
30 criterion g) is the covering markers are substantially evenly distributed across a chromosome or a  
31 chromosomal segment, wherein the average chromosomal intermarker distance of the covering markers  
32 is greater than 2 cM or the equivalent thereof, and wherein the conditional probability the covering  
33 markers were chosen essentially randomly from substantially the known set of bi-allelic markers with  
34 least common allele frequencies between 0.3 inclusive and 0.5 inclusive is less than or equal to 10  
35 percent, wherein the conditional probability is substantially conditional on (1) the approximate

PCT/US99/04376

PCT/US 99/04376

IPEAUS 22 MAY 2000

56  
-82

1 chromosomal distribution of the covering markers, (2) the marker type of each covering marker and (3)  
 2 the CL-F region being N covered to within the CL-F distance  $\delta$  by the two or more bi-allelic covering  
 3 markers;

4  
 5 criterion h) is the covering markers are substantially evenly distributed across a chromosome or a  
 6 chromosomal segment, wherein the average chromosomal intermarker distance of the covering markers  
 7 is less than or equal to 2 cM or the equivalent thereof, and wherein collection C is essentially the  
 8 collection of known groups of bi-allelic markers with least common allele frequencies between 0.2  
 9 inclusive and 0.5 inclusive, each group in the collection being substantially similar to the covering  
 10 markers as a group, wherein a group of bi-allelic markers is a member of collection C if and only if the  
 11 group substantially meets criteria (1), (2), (3) and (4); wherein criterion (1) is, each marker in the group  
 12 is chosen from substantially the known set of bi-allelic markers with least common allele frequencies  
 13 between 0.2 inclusive and 0.5 inclusive, wherein criterion (2) is, the number of markers in the group is  
 14 the same as the number of covering markers, wherein criterion (3) is, the chromosomal distribution of  
 15 the group of markers and the covering markers is substantially similar, and wherein criterion (4) is, the  
 16 marker type of each group marker and the covering marker with substantially the same chromosomal  
 17 location is the same, wherein a group that is a member of collection C meets criterion (5) if and only if  
 18 (5) the CL-F region is N covered to within a CL-F distance  $\delta$  by the two or more bi-allelic covering  
 19 markers, wherein P is essentially the proportion of groups in collection C that meet criterion (5); wherein  
 20 P is less than 90 percent;

21  
 22 and criterion i) is the covering markers are substantially evenly distributed across a chromosome or a  
 23 chromosomal segment, wherein the average chromosomal intermarker distance of the covering markers  
 24 is greater than 2 cM or the equivalent thereof, and wherein collection C is essentially the collection of  
 25 known groups of bi-allelic markers with least common allele frequencies between 0.3 inclusive and 0.5  
 26 inclusive, each group in the collection being substantially similar to the covering markers as a group;  
 27 wherein a group of bi-allelic markers is a member of collection C if and only if the group substantially  
 28 meets criteria (1), (2), (3) and (4); wherein criterion (1) is, each marker in the group is chosen from  
 29 substantially the known set of bi-allelic markers with least common allele frequencies between 0.3  
 30 inclusive and 0.5 inclusive, wherein criterion (2) is, the number of markers in the group is the same as  
 31 the number of covering markers, wherein criterion (3) is, the chromosomal distribution of the group of  
 32 markers and the covering markers is substantially similar, and wherein criterion (4) is, the marker type of  
 33 each group marker and the covering marker with substantially the same chromosomal location is the  
 34 same; wherein a group that is a member of collection C meets criterion (5) if and only if (5) the CL-F  
 35 region is N covered to within a CL-F distance  $\delta$  by the two or more bi-allelic covering markers; wherein P

**AMENDED SHEET**

●CT/US 99/04376

IPEA/US 22 MAY 2000

PCT/US99/04376

57

-83

- 1 is essentially the proportion of groups in collection C that meet criterion (5); wherein P is less than 90
- 2 percent.
- 3 19. A process as in claim 18, wherein  $\delta$  is less than or equal to about [1 cM, 0.15] or the equivalent
- 4 thereof
- 5 ~~20. A process as in any one of claims 3-19, wherein there is a subgroup of the covering markers, and~~
- 6 ~~the markers in the subgroup are a majority of the covering markers, and each marker in the subgroup is~~
- 7 ~~an SNP, or a bi-allelic marker equivalent formed only from one or more SNPs.~~
- 8 21. A process as in claim 20, wherein the process comprises a computer program
- 9 22. A process as in claim 20 wherein  $L_{MC}$  is less than or equal to about 250,000 bp or the equivalent
- 10 thereof,  $W_{MC}$  is less than or equal to about 0.15, wherein the species is human being, wherein the same
- 11 statistical linkage test based on allelic association is chosen for each covering marker in step b).
- 12 ~~23. A process as in any one of claim 22, wherein the process comprises a computer program~~
- 13 24. An apparatus for identifying bi-allelic markers linked to a bi-allelic genetic characteristic gene in a
- 14 species of creatures, comprising. means to practice each of the steps of a process as in claim 20.
- 15 25. An apparatus as in claim 24, wherein the apparatus comprises oligonucleotide technology or mass
- 16 spectrometry.
- 17 26. An apparatus for identifying bi-allelic markers linked to a bi-allelic genetic characteristic gene in a
- 18 species of creatures, comprising. means to practice each of the steps of a process as in any one of the
- 19 claims 3-19.
- 20 27. An apparatus as in claim 26, wherein the apparatus comprises a computer, the computer being
- 21 supplied with proper data and instructions
- 22
- 23 28. A process for localizing a bi-allelic genetic characteristic gene in a species of creatures to a CL-F
- 24 location, comprising the steps of. a process as in claim 20, further comprising the step f) of. localizing
- 25 the gene to the CL-F location of one or more markers that show evidence for linkage based on the
- 26 calculations of step e).
- 27 29. A process for localizing a bi-allelic genetic characteristic gene in a species of creatures to a CL-F
- 28 location, comprising the steps of. a process in any one of the claims 3-19; further comprising. the step f)
- 29 of. localizing the gene to the CL-F location of one or more markers that show evidence for linkage
- 30 based on the calculations of step e)
- 31 30. A process as in claim 29, wherein the process comprises a computer program
- 32 31. An apparatus for localizing a bi-allelic genetic characteristic gene in a species of creatures to a CL-F
- 33 location, comprising. means to practice each of the steps of a process as in claim 29

AMENDED SHEET

58  
~~84~~

32. An apparatus for localizing a bi-allelic genetic characteristic gene in a species of creatures to a CL-F location as in claim 31, wherein the apparatus comprises a computer, the computer being supplied with proper data and instructions.

33. A process for obtaining genotype data/sample allele frequency data for each bi-allelic marker of a group of two or more bi-allelic covering markers in the chromosomal DNA of one or more individuals of a sample, each individual in the sample being a member of the same species, comprising

a) determining information on the presence or absence of each allele of each bi-allelic marker of a group of two or more bi-allelic covering markers in the chromosomal DNA of one or more individuals of a sample, a CL-F region being systematically covered by the two or more bi-allelic covering markers, the CL-F region being a collection of points on a two-dimensional plane, the two-dimensional plane having the two orthogonal dimensions of chromosomal location and least common allele frequency, and

b) transforming the information of step a) into genotype data/sample allele frequency data for each marker of the group

34. A process for obtaining genotype data/sample allele frequency data as in claim 33, wherein the CL-F region is N covered to within a CL-F distance  $\delta$  by the two or more bi-allelic covering markers, so that each point in the region is within the CL-F distance  $\delta$  of N or more of the covering markers, wherein  $\delta$  is equal to about  $[\delta_{CL}, 0.25]$  or the equivalent thereof,  $\delta_{CL}$  is equal to the largest chromosomal length, computed by any method, for which linkage disequilibrium has been observed between any polymorphisms in any population of the species, N is an integer greater than or equal to 1.

35. A process for obtaining genotype data/sample allele frequency data as in claim 34, wherein the CL-F region includes 81 percent or more of the centerpoints of the matrix centerpoint lattice of a CL-F matrix, the number of cells in the matrix being greater than or equal to three, wherein the matrix has R rows and C columns, each cell of the matrix being of length  $L_{MC}$  and width  $W_{MC}$ , and  $L_{MC}$  being less than or equal to about  $2\delta_{CL}$ , and  $W_{MC}$  being less than or equal to 0.5,  $\delta_{CL}$  is equal to the largest chromosomal length, computed by any method, for which linkage disequilibrium has been observed between any polymorphisms in any population of the species, there being N or more covering markers in each cell of 81 percent or more of the cells of the matrix, N is an integer greater than or equal to 1; the covering markers being distributed over a chromosomal region of interest, the region of interest being approximately the smallest chromosome interval that contains all of the covering markers, and the covering markers comprising essentially less than all of the polymorphisms in the region of interest

59  
85

41. A process for obtaining genotype data/sample allele frequency data as in claim 35, wherein the covering markers are substantially evenly distributed across a chromosome or a chromosomal segment, wherein the average chromosomal intermarker distance of the covering markers is greater than 2 cM or the equivalent thereof, and wherein the conditional probability the covering markers were chosen essentially randomly from substantially the known set of bi-allelic markers with least common allele

**AMENDED SHEET**

PCT/US99/04376

40  
86

1 frequencies between 0.3 inclusive and 0.5 inclusive is less than or equal to 10 percent; wherein the  
2 conditional probability is substantially conditional on (1) the approximate chromosomal distribution of the  
3 covering markers, (2) the marker type of each covering marker and (3) there being N or more covering  
4 markers in each cell of 81 percent or more of the cells of the matrix

5 42. A process for obtaining genotype data/sample allele frequency data as in claim 40, wherein the  
6 chromosome or the chromosomal segment consists essentially of a set of nonoverlapping chromosome  
7 segments of substantially equal length, and wherein each chromosome segment of the set has two or  
8 less covering markers located thereon, wherein one and only one covering marker is located on each of  
9 80 percent or more of the chromosome segments of the set, and wherein zero or two and only two  
10 covering markers are located on each of 20 percent or less of the chromosome segments of the set,  
11 and wherein each chromosome segment that borders a chromosome segment with zero covering  
12 markers located thereon has only one or two covering markers located thereon, and wherein each  
13 chromosome segment that borders a chromosome segment with two covering markers located thereon  
14 has only one or zero covering markers located thereon; and wherein the conditional probability the  
15 covering markers were chosen essentially randomly from substantially the known set of bi-allelic  
16 markers with least common allele frequencies between 0.2 inclusive and 0.5 inclusive is less than or  
17 equal to 10 percent; wherein the conditional probability is substantially conditional on (1) the  
18 chromosomal distribution of the covering markers on the chromosome segments of the set, (2) the  
19 marker type of each covering marker and (3) there being N or more covering markers in each cell of 81  
20 percent or more of the cells of the matrix.

21 43. A process for obtaining genotype data/sample allele frequency data as in claim 41, wherein the  
22 chromosome or the chromosomal segment consists essentially of a set of nonoverlapping chromosome  
23 segments of substantially equal length, and wherein each chromosome segment of the set has two or  
24 less covering markers located thereon, wherein one and only one covering marker is located on each of  
25 80 percent or more of the chromosome segments of the set, and wherein zero or two and only two  
26 covering markers are located on each of 20 percent or less of the chromosome segments of the set,  
27 and wherein each chromosome segment that borders a chromosome segment with zero covering  
28 markers located thereon has only one or two covering markers located thereon, and wherein each  
29 chromosome segment that borders a chromosome segment with two covering markers located thereon  
30 has only one or zero covering markers located thereon; and wherein the conditional probability the  
31 covering markers were chosen essentially randomly from substantially the known set of bi-allelic  
32 markers with least common allele frequencies between 0.3 inclusive and 0.5 inclusive is less than or  
33 equal to 10 percent; wherein the conditional probability is substantially conditional on (1) the  
34 chromosomal distribution of the covering markers on the chromosome segments of the set, (2) the

**AMENDED SHEET**

61  
87

marker type of each covering marker and (3) there being N or more covering markers in each cell of 81 percent or more of the cells of the matrix

44. A process for obtaining genotype data/sample allele frequency data as in claim 35, wherein the covering markers are substantially evenly distributed across a chromosome or a chromosomal segment, wherein the average chromosomal intermarker distance of the covering markers is less than or equal to 2 cM or the equivalent thereof, and wherein collection C is essentially the collection of known groups of bi-allelic markers with least common allele frequencies between 0.2 inclusive and 0.5 inclusive, each group in the collection being substantially similar to the covering markers as a group; wherein a group of bi-allelic markers is a member of collection C if and only if the group substantially meets criteria (1), (2), (3) and (4), wherein criterion (1) is, each marker in the group is chosen from substantially the known set of bi-allelic markers with least common allele frequencies between 0.2 inclusive and 0.5 inclusive, wherein criterion (2) is, the number of markers in the group is the same as the number of covering markers, wherein criterion (3) is, the chromosomal distribution of the group of markers and the covering markers is substantially similar, and wherein criterion (4) is, the marker type of each group marker and the covering marker with substantially the same chromosomal location is the same; wherein a group that is a member of collection C meets criterion (5) if and only if (5) there are N or more of the group markers in each cell of 81 percent or more of the cells of a CL-F matrix with cells having length  $L_{MC}$  and width  $W_{MC}$  in R rows and C columns; wherein P is essentially the proportion of groups in collection C that meet criterion (5); wherein P is less than 90 percent.

45. A process for obtaining genotype data/sample allele frequency data as in claim 35, wherein the covering markers are substantially evenly distributed across a chromosome or a chromosomal segment, wherein the average chromosomal intermarker distance of the covering markers is greater than 2 cM or the equivalent thereof, and wherein collection C is essentially the collection of known groups of bi-allelic markers with least common allele frequencies between 0.3 inclusive and 0.5 inclusive, each group in the collection being substantially similar to the covering markers as a group, wherein a group of bi-allelic markers is a member of collection C if and only if the group substantially meets criteria (1), (2), (3) and (4); wherein criterion (1) is, each marker in the group is chosen from substantially the known set of bi-allelic markers with least common allele frequencies between 0.3 inclusive and 0.5 inclusive, wherein criterion (2) is, the number of markers in the group is the same as the number of covering markers, wherein criterion (3) is, the chromosomal distribution of the group of markers and the covering markers is substantially similar, and wherein criterion (4) is, the marker type of each group marker and the covering marker with substantially the same chromosomal location is the same, wherein a group that is a member of collection C meets criterion (5) if and only if (5) there are N or more of the group markers in each cell of 81 percent or more of the cells of a CL-F matrix with cells having length  $L_{MC}$  and width  $W_{MC}$

**AMENDED SHEET**

PCT/US99/04376

62  
88

1 in R rows and C columns; wherein P is essentially the proportion of groups in collection C that meet  
2 criterion (5); wherein P is less than 90 percent

3 46. A process for obtaining genotype data/sample allele frequency data as in claim 35, wherein the  
4 covering markers are substantially evenly distributed across a chromosome or a chromosomal segment,  
5 wherein the average chromosomal intermarker distance of the covering markers is less than or equal to  
6 2 cM or the equivalent thereof; wherein the chromosome or the chromosomal segment consists  
7 essentially of a set of nonoverlapping chromosome segments of substantially equal length, and wherein  
8 each chromosome segment of the set has two or less covering markers located thereon, wherein one  
9 and only one covering marker is located on each of 80 percent or more of the chromosome segments of  
10 the set, and wherein zero or two and only two covering markers are located on each of 20 percent or  
11 less of the chromosome segments of the set, and wherein each chromosome segment that borders a  
12 chromosome segment with zero covering markers located thereon has only one or two covering  
13 markers located thereon, and wherein each chromosome segment that borders a chromosome segment  
14 with two covering markers located thereon has only one or zero covering markers located thereon,  
15 wherein collection D is essentially the collection of known groups of bi-allelic markers with least common  
16 allele frequencies between 0.2 inclusive and 0.5 inclusive, each group in the collection being  
17 substantially similar to the covering markers as a group; wherein a group of bi-allelic markers is a  
18 member of collection D if and only if the group substantially meets criteria (1), (2), and (3) (1) each  
19 marker in the group is chosen from substantially the known set of bi-allelic markers with least common  
20 allele frequencies between 0.2 inclusive and 0.5 inclusive, (2) the number of covering markers and the  
21 number of group markers located on each chromosome segment of the set is the same, and (3) there is  
22 a group marker of the same type as each covering marker located on the same chromosome segment  
23 of the set as each covering marker; wherein a group that is a member of collection D substantially  
24 meets criterion (5) if and only if (5) there are N or more of the group markers in each cell of the matrix,  
25 wherein P is essentially the proportion of groups in collection D that meet criterion (5), wherein P is less  
26 than 90 percent.

27 47. A process for obtaining genotype data/sample allele frequency data as in claim 35, wherein the  
28 covering markers are substantially evenly distributed across a chromosome or a chromosomal segment,  
29 wherein the average chromosomal intermarker distance of the covering markers is greater than 2 cM or  
30 the equivalent thereof, wherein the chromosome or the chromosomal segment consists essentially of a  
31 set of nonoverlapping chromosome segments of substantially equal length, and wherein each  
32 chromosome segment of the set has two or less covering markers located thereon, wherein one and  
33 only one covering marker is located on each of 80 percent or more of the chromosome segments of the  
34 set, and wherein zero or two and only two covering markers are located on each of 20 percent or less of  
35 the chromosome segments of the set, and wherein each chromosome segment that borders a

AMENDED SHEET



43  
-89-

1 chromosome segment with zero covering markers located thereon has only one or two covering  
2 markers located thereon, and wherein each chromosome segment that borders a chromosome segment  
3 with two covering markers located thereon has only one or zero covering markers located thereon,  
4 wherein collection D is essentially the collection of known groups of bi-allelic markers with least common  
5 allele frequencies between 0.3 inclusive and 0.5 inclusive, each group in the collection being  
6 substantially similar to the covering markers as a group, wherein a group of bi-allelic markers is a  
7 member of collection D if and only if the group substantially meets criteria (1), (2), and (3). (1) each  
8 marker in the group is chosen from substantially the known set of bi-allelic markers with least common  
9 allele frequencies between 0.3 inclusive and 0.5 inclusive, (2) the number of covering markers and the  
10 number of group markers located on each chromosome segment of the set is the same, and (3) there is  
11 a group marker of the same type as each covering marker located on the same chromosome segment  
12 of the set as each covering marker; wherein a group that is a member of collection D substantially  
13 meets criterion (5) if and only if (5) there are N or more of the group markers in each cell of the matrix;  
14 wherein P is essentially the proportion of groups in collection D that meet criterion (5), wherein P is less  
15 than 90 percent

16 48. A process for obtaining genotype data/sample allele frequency data as in claim 34, wherein the  
17 covering markers meet criterion a) and one or more of the criteria b), c), d), e), f), g), h) or i), wherein  
18

19 criterion a) is the covering markers are distributed over a chromosomal region of interest, the region of  
20 interest being approximately the smallest chromosome interval that contains all of the covering markers,  
21 and the covering markers comprise essentially less than all of the polymorphisms in the region of  
22 interest;

23  
24 criterion b) is the covering markers are substantially nonevenly distributed across a chromosome or a  
25 chromosomal segment;

26  
27 criterion c) is the covering markers are substantially evenly distributed across a chromosome or a  
28 chromosomal segment, and there is a subgroup of one or more of the covering markers, and each of  
29 the markers in the subgroup is chosen without substantial preference for the least common allele  
30 frequency of each of the markers in the subgroup being close to 0.5, and the number of covering  
31 markers in the subgroup is about 5 percent or more of the total number of covering markers;

32  
33 criterion d) is the covering markers are substantially evenly distributed across a chromosome or a  
34 chromosomal segment, and there is a subgroup of one or more of the covering markers, and each of  
35 the markers in the subgroup is chosen without substantial preference for the least common allele

**AMENDED SHEET**

PCT/US99/04376

64  
90

frequency of each of the markers in the subgroup being close to 0.5,

criterion e) is the covering markers are substantially evenly distributed across a chromosome or a chromosomal segment, wherein (1) the average chromosomal intermarker distance of the covering markers is greater than 2 cM or the equivalent thereof and the least common allele frequency of one or more of the covering markers is less than 0.3, or wherein (2) the least common allele frequency of one or more of the covering markers is less than 0.2, or wherein (3) the average chromosomal intermarker distance of the covering markers is greater than 10 cM or the equivalent thereof,

criterion f) is the covering markers are substantially evenly distributed across a chromosome or a chromosomal segment, wherein the average chromosomal intermarker distance of the covering markers is less than or equal to 2 cM or the equivalent thereof, and wherein the conditional probability the covering markers were chosen essentially randomly from substantially the known set of bi-allelic markers with least common allele frequencies between 0.2 inclusive and 0.5 inclusive is less than or equal to 10 percent; wherein the conditional probability is substantially conditional on (1) the approximate chromosomal distribution of the covering markers, (2) the marker type of each covering marker and (3) the CL-F region being N covered to within the CL-F distance  $\delta$  by the two or more bi-allelic covering markers,

criterion g) is the covering markers are substantially evenly distributed across a chromosome or a chromosomal segment, wherein the average chromosomal intermarker distance of the covering markers is greater than 2 cM or the equivalent thereof; and wherein the conditional probability the covering markers were chosen essentially randomly from substantially the known set of bi-allelic markers with least common allele frequencies between 0.3 inclusive and 0.5 inclusive is less than or equal to 10 percent; wherein the conditional probability is substantially conditional on (1) the approximate chromosomal distribution of the covering markers, (2) the marker type of each covering marker and (3) the CL-F region being N covered to within the CL-F distance  $\delta$  by the two or more bi-allelic covering markers,

criterion h) is the covering markers are substantially evenly distributed across a chromosome or a chromosomal segment, wherein the average chromosomal intermarker distance of the covering markers is less than or equal to 2 cM or the equivalent thereof; and wherein collection C is essentially the collection of known groups of bi-allelic markers with least common allele frequencies between 0.2 inclusive and 0.5 inclusive, each group in the collection being substantially similar to the covering markers as a group, wherein a group of bi-allelic markers is a member of collection C if and only if the

**AMENDED SHEET**

PCT/US99/04376

65  
91

group substantially meets criteria (1), (2), (3) and (4), wherein criterion (1) is, each marker in the group is chosen from substantially the known set of bi-allelic markers with least common allele frequencies between 0.2 inclusive and 0.5 inclusive, wherein criterion (2) is, the number of markers in the group is the same as the number of covering markers, wherein criterion (3) is, the chromosomal distribution of the group of markers and the covering markers is substantially similar, and wherein criterion (4) is, the marker type of each group marker and the covering marker with substantially the same chromosomal location is the same; wherein a group that is a member of collection C meets criterion (5) if and only if (5) the CL-F region is N covered to within a CL-F distance  $\delta$  by the two or more bi-allelic covering markers; wherein P is essentially the proportion of groups in collection C that meet criterion (5); wherein P is less than 90 percent;

and criterion i) is the covering markers are substantially evenly distributed across a chromosome or a chromosomal segment, wherein the average chromosomal intermarker distance of the covering markers is greater than 2 cM or the equivalent thereof, and wherein collection C is essentially the collection of known groups of bi-allelic markers with least common allele frequencies between 0.3 inclusive and 0.5 inclusive, each group in the collection being substantially similar to the covering markers as a group; wherein a group of bi-allelic markers is a member of collection C if and only if the group substantially meets criteria (1), (2), (3) and (4); wherein criterion (1) is, each marker in the group is chosen from substantially the known set of bi-allelic markers with least common allele frequencies between 0.3 inclusive and 0.5 inclusive, wherein criterion (2) is, the number of markers in the group is the same as the number of covering markers, wherein criterion (3) is, the chromosomal distribution of the group of markers and the covering markers is substantially similar, and wherein criterion (4) is, the marker type of each group marker and the covering marker with substantially the same chromosomal location is the same, wherein a group that is a member of collection C meets criterion (5) if and only if (5) the CL-F region is N covered to within a CL-F distance  $\delta$  by the two or more bi-allelic covering markers; wherein P is essentially the proportion of groups in collection C that meet criterion (5), wherein P is less than 90 percent

49. A process for obtaining genotype data/sample allele frequency data as in claim 48, wherein  $\delta$  is less than or equal to about [1 cM, 0.15] or the equivalent thereof.

~~50. A process for obtaining genotype data/sample allele frequency data as in any one of claims 33-49, wherein there is a subgroup of the covering markers, and the markers in the subgroup are a majority of the covering markers, and each marker in the subgroup is an SNP, or a bi-allelic marker equivalent formed only from one or more SNPs.~~

51. A process for obtaining genotype data/sample allele frequency data as in claim 50, wherein the process comprises a computer program

AMENDED SHEET

66  
92

52. A process for obtaining genotype data/sample allele frequency data as in claim 50 wherein  $L_{MC}$  is less than or equal to about 250,000 bp or the equivalent thereof,  $W_{MC}$  is less than or equal to about 0.15, wherein the species is human being.

53. A process for obtaining genotype data/sample allele frequency data as in claim 52, wherein the process comprises a computer program.

54. An apparatus for obtaining genotype data/sample allele frequency data for each bi-allelic marker of a group of two or more bi-allelic covering markers in the chromosomal DNA of one or more individuals of a sample, each individual in the sample being a member of the same species, comprising means to practice each of the steps of a process as in any one of the claims 33-49.

55. An apparatus for obtaining genotype data/sample allele frequency data for each bi-allelic marker of a group of two or more bi-allelic covering markers in the chromosomal DNA of one or more individuals of a sample, each individual in the sample being a member of the same species, comprising means to practice each of the steps of a process as in claim 50.

56. An apparatus as in claim 55, wherein the apparatus comprises oligonucleotide technology or mass spectrometry

57. An apparatus as in claim 54, wherein the apparatus comprises oligonucleotide technology or mass spectrometry.

58. An apparatus as in claim 54, wherein the apparatus comprises a computer, the computer being supplied with proper data and instructions.

59. A use of one or more copies of a set of oligonucleotides to determine genotype data/sample allele frequency data for each bi-allelic marker of a group of two or more bi-allelic covering markers for one or more individuals, each individual being a member of the same species, wherein the group of covering markers systematically cover a CL-F region, the CL-F region being a collection of points on a two-dimensional plane, the two-dimensional plane having the two orthogonal dimensions of chromosomal location and least common allele frequency

60. A use as in claim 59, wherein the CL-F region is N covered to within a CL-F distance  $\delta$  by the two or more bi-allelic covering markers, so that each point in the region is within the CL-F distance  $\delta$  of N or more of the covering markers, wherein  $\delta$  is equal to about  $[\delta_{CL}, 0.25]$  or the equivalent thereof,  $\delta_{CL}$  is equal to the largest chromosomal length, computed by any method, for which linkage disequilibrium has been observed between any polymorphisms in any population of the species, N is an integer greater than or equal to 1.

61. A use as in claim 59, wherein the CL-F region includes 81 percent or more of the centerpoints of the matrix centerpoint lattice of a CL-F matrix, the number of cells in the matrix being greater than or equal

**AMENDED SHEET**

67  
93

1 to three, wherein the matrix has R rows and C columns, each cell of the matrix being of length  $L_{MC}$  and  
2 width  $W_{MC}$ , and  $L_{MC}$  being less than or equal to about  $2\delta_{CL}$ , and  $W_{MC}$  being less than or equal to 0.5,  $\delta_{CL}$   
3 is equal to the largest chromosomal length, computed by any method, for which linkage disequilibrium  
4 has been observed between any polymorphisms in any population of the species, there being N or more  
5 covering markers in each cell of 81 percent or more of the cells of the matrix, N is an integer greater  
6 than or equal to 1, the covering markers being distributed over a chromosomal region of interest, the  
7 region of interest being approximately the smallest chromosome interval that contains all of the covering  
8 markers, and the covering markers comprising essentially less than all of the polymorphisms in the  
9 region of interest

10 62. A use as in claim 61, wherein the covering markers are substantially nonevenly distributed across a  
11 chromosome or a chromosomal segment.

12 63. A use as in claim 61, wherein the covering markers are substantially evenly distributed across a  
13 chromosome or a chromosomal segment, and wherein there is a subgroup of one or more of the  
14 covering markers, and each of the markers in the subgroup is chosen without substantial preference for  
15 the least common allele frequency of each of the markers in the subgroup being close to 0.5, and the  
16 number of covering markers in the subgroup is about 5 percent or more of the total number of covering  
17 markers.

18 64. A use as in claim 61, wherein the covering markers are substantially evenly distributed across a  
19 chromosome or a chromosomal segment, and wherein there is a subgroup of one or more of the  
20 covering markers, and each of the markers in the subgroup is chosen without substantial preference for  
21 the least common allele frequency of each of the markers in the subgroup being close to 0.5.

22 65. A use as in claim 61, wherein the covering markers are substantially evenly distributed across a  
23 chromosome or a chromosomal segment, wherein (1) the average chromosomal intermarker distance of  
24 the covering markers is greater than 2 cM or the equivalent thereof and the least common allele  
25 frequency of one or more of the covering markers is less than 0.3, or wherein (2) the least common  
26 allele frequency of one or more of the covering markers is less than 0.2, or wherein (3) the average  
27 chromosomal intermarker distance of the covering markers is greater than 10 cM or the equivalent  
28 thereof

29 66. A use as in claim 61, wherein the covering markers are substantially evenly distributed across a  
30 chromosome or a chromosomal segment, wherein the average chromosomal intermarker distance of  
31 the covering markers is less than or equal to 2 cM or the equivalent thereof, and wherein the conditional  
32 probability the covering markers were chosen essentially randomly from substantially the known set of  
33 bi-allelic markers with least common allele frequencies between 0.2 inclusive and 0.5 inclusive is less  
34 than or equal to 10 percent; wherein the conditional probability is substantially conditional on (1) the  
35 approximate chromosomal distribution of the covering markers, (2) the marker type of each covering

**AMENDED SHEET**

48

94

1 marker and (3) there being N or more covering markers in each cell of 81 percent or more of the cells of  
2 the matrix.

3 67. A use as in claim 61, wherein the covering markers are substantially evenly distributed across a  
4 chromosome or a chromosomal segment, wherein the average chromosomal intermarker distance of  
5 the covering markers is greater than 2 cM or the equivalent thereof, and wherein the conditional  
6 probability the covering markers were chosen essentially randomly from substantially the known set of  
7 bi-allelic markers with least common allele frequencies between 0.3 inclusive and 0.5 inclusive is less  
8 than or equal to 10 percent; wherein the conditional probability is substantially conditional on (1) the  
9 approximate chromosomal distribution of the covering markers, (2) the marker type of each covering  
10 marker and (3) there being N or more covering markers in each cell of 81 percent or more of the cells of  
11 the matrix.

12 68. A use as in claim 66, wherein the chromosome or the chromosomal segment consists essentially of  
13 a set of nonoverlapping chromosome segments of substantially equal length, and wherein each  
14 chromosome segment of the set has two or less covering markers located thereon, wherein one and  
15 only one covering marker is located on each of 80 percent or more of the chromosome segments of the  
16 set, and wherein zero or two and only two covering markers are located on each of 20 percent or less of  
17 the chromosome segments of the set, and wherein each chromosome segment that borders a  
18 chromosome segment with zero covering markers located thereon has only one or two covering  
19 markers located thereon, and wherein each chromosome segment that borders a chromosome segment  
20 with two covering markers located thereon has only one or zero covering markers located thereon; and  
21 wherein the conditional probability the covering markers were chosen essentially randomly from  
22 substantially the known set of bi-allelic markers with least common allele frequencies between 0.2  
23 inclusive and 0.5 inclusive is less than or equal to 10 percent, wherein the conditional probability is  
24 substantially conditional on (1) the chromosomal distribution of the covering markers on the  
25 chromosome segments of the set, (2) the marker type of each covering marker and (3) there being N or  
26 more covering markers in each cell of 81 percent or more of the cells of the matrix.

27 69. A use as in claim 67, wherein the chromosome or the chromosomal segment consists essentially of  
28 a set of nonoverlapping chromosome segments of substantially equal length, and wherein each  
29 chromosome segment of the set has two or less covering markers located thereon, wherein one and  
30 only one covering marker is located on each of 80 percent or more of the chromosome segments of the  
31 set, and wherein zero or two and only two covering markers are located on each of 20 percent or less of  
32 the chromosome segments of the set, and wherein each chromosome segment that borders a  
33 chromosome segment with zero covering markers located thereon has only one or two covering  
34 markers located thereon, and wherein each chromosome segment that borders a chromosome segment  
35 with two covering markers located thereon has only one or zero covering markers located thereon; and

**AMENDED SHEET**

69  
85

wherein the conditional probability the covering markers were chosen essentially randomly from substantially the known set of bi-allelic markers with least common allele frequencies between 0.3 inclusive and 0.5 inclusive is less than or equal to 10 percent; wherein the conditional probability is substantially conditional on (1) the chromosomal distribution of the covering markers on the chromosome segments of the set, (2) the marker type of each covering marker and (3) there being N or more covering markers in each cell of 81 percent or more of the cells of the matrix

70. A use as in claim 61, wherein the covering markers are substantially evenly distributed across a chromosome or a chromosomal segment, wherein the average chromosomal intermarker distance of the covering markers is less than or equal to 2 cM or the equivalent thereof; and wherein collection C is essentially the collection of known groups of bi-allelic markers with least common allele frequencies between 0.2 inclusive and 0.5 inclusive, each group in the collection being substantially similar to the covering markers as a group; wherein a group of bi-allelic markers is a member of collection C if and only if the group substantially meets criteria (1), (2), (3) and (4); wherein criterion (1) is, each marker in the group is chosen from substantially the known set of bi-allelic markers with least common allele frequencies between 0.2 inclusive and 0.5 inclusive, wherein criterion (2) is, the number of markers in the group is the same as the number of covering markers, wherein criterion (3) is, the chromosomal distribution of the group of markers and the covering markers is substantially similar, and wherein criterion (4) is, the marker type of each group marker and the covering marker with substantially the same chromosomal location is the same; wherein a group that is a member of collection C meets criterion (5) if and only if (5) there are N or more of the group markers in each cell of 81 percent or more of the cells of a CL-F matrix with cells having length  $L_{MC}$  and width  $W_{MC}$  in R rows and C columns, wherein P is essentially the proportion of groups in collection C that meet criterion (5); wherein P is less than 90 percent

71. A use as in claim 61, wherein the covering markers are substantially evenly distributed across a chromosome or a chromosomal segment, wherein the average chromosomal intermarker distance of the covering markers is greater than 2 cM or the equivalent thereof, and wherein collection C is essentially the collection of known groups of bi-allelic markers with least common allele frequencies between 0.3 inclusive and 0.5 inclusive, each group in the collection being substantially similar to the covering markers as a group, wherein a group of bi-allelic markers is a member of collection C if and only if the group substantially meets criteria (1), (2), (3) and (4), wherein criterion (1) is, each marker in the group is chosen from substantially the known set of bi-allelic markers with least common allele frequencies between 0.3 inclusive and 0.5 inclusive, wherein criterion (2) is, the number of markers in the group is the same as the number of covering markers, wherein criterion (3) is, the chromosomal distribution of the group of markers and the covering markers is substantially similar, and wherein criterion (4) is, the marker type of each group marker and the covering marker with substantially the

**AMENDED SHEET**

70  
96

1 same chromosomal location is the same; wherein a group that is a member of collection C meets  
2 criterion (5) if and only if (5) there are N or more of the group markers in each cell of 81 percent or more  
3 of the cells of a CL-F matrix with cells having length  $L_{MC}$  and width  $W_{MC}$  in R rows and C columns;  
4 wherein P is essentially the proportion of groups in collection C that meet criterion (5); wherein P is less  
5 than 90 percent

6 72. A use as in claim 61, wherein the covering markers are substantially evenly distributed across a  
7 chromosome or a chromosomal segment, wherein the average chromosomal intermarker distance of  
8 the covering markers is less than or equal to 2 cM or the equivalent thereof; wherein the chromosome or  
9 the chromosomal segment consists essentially of a set of nonoverlapping chromosome segments of  
10 substantially equal length, and wherein each chromosome segment of the set has two or less covering  
11 markers located thereon, wherein one and only one covering marker is located on each of 80 percent or  
12 more of the chromosome segments of the set, and wherein zero or two and only two covering markers  
13 are located on each of 20 percent or less of the chromosome segments of the set, and wherein each  
14 chromosome segment that borders a chromosome segment with zero covering markers located thereon  
15 has only one or two covering markers located thereon, and wherein each chromosome segment that  
16 borders a chromosome segment with two covering markers located thereon has only one or zero  
17 covering markers located thereon; wherein collection D is essentially the collection of known groups of  
18 bi-allelic markers with least common allele frequencies between 0.2 inclusive and 0.5 inclusive, each  
19 group in the collection being substantially similar to the covering markers as a group; wherein a group of  
20 bi-allelic markers is a member of collection D if and only if the group substantially meets criteria (1), (2),  
21 and (3): (1) each marker in the group is chosen from substantially the known set of bi-allelic markers  
22 with least common allele frequencies between 0.2 inclusive and 0.5 inclusive, (2) the number of  
23 covering markers and the number of group markers located on each chromosome segment of the set is  
24 the same, and (3) there is a group marker of the same type as each covering marker located on the  
25 same chromosome segment of the set as each covering marker; wherein a group that is a member of  
26 collection D substantially meets criterion (5) if and only if (5) there are N or more of the group markers in  
27 each cell of the matrix, wherein P is essentially the proportion of groups in collection D that meet  
28 criterion (5); wherein P is less than 90 percent

29 73. A use as in claim 61, wherein the covering markers are substantially evenly distributed across a  
30 chromosome or a chromosomal segment, wherein the average chromosomal intermarker distance of  
31 the covering markers is greater than 2 cM or the equivalent thereof; wherein the chromosome or the  
32 chromosomal segment consists essentially of a set of nonoverlapping chromosome segments of  
33 substantially equal length, and wherein each chromosome segment of the set has two or less covering  
34 markers located thereon, wherein one and only one covering marker is located on each of 80 percent or  
35 more of the chromosome segments of the set, and wherein zero or two and only two covering markers



-71  
97

1 are located on each of 20 percent or less of the chromosome segments of the set, and wherein each  
2 chromosome segment that borders a chromosome segment with zero covering markers located thereon  
3 has only one or two covering markers located thereon, and wherein each chromosome segment that  
4 borders a chromosome segment with two covering markers located thereon has only one or zero  
5 covering markers located thereon; wherein collection D is essentially the collection of known groups of  
6 bi-allelic markers with least common allele frequencies between 0.3 inclusive and 0.5 inclusive, each  
7 group in the collection being substantially similar to the covering markers as a group, wherein a group of  
8 bi-allelic markers is a member of collection D if and only if the group substantially meets criteria (1), (2),  
9 and (3): (1) each marker in the group is chosen from substantially the known set of bi-allelic markers  
10 with least common allele frequencies between 0.3 inclusive and 0.5 inclusive, (2) the number of  
11 covering markers and the number of group markers located on each chromosome segment of the set is  
12 the same, and (3) there is a group marker of the same type as each covering marker located on the  
13 same chromosome segment of the set as each covering marker; wherein a group that is a member of  
14 collection D substantially meets criterion (5) if and only if (5) there are N or more of the group markers in  
15 each cell of the matrix; wherein P is essentially the proportion of groups in collection D that meet  
16 criterion (5); wherein P is less than 90 percent

17 74. A use as in claim 60, wherein the covering markers meet criterion a) and one or more of the criteria  
18 b), c), d), e), f), g), h) or i), wherein

19  
20 criterion a) is the covering markers are distributed over a chromosomal region of interest, the region of  
21 interest being approximately the smallest chromosome interval that contains all of the covering markers,  
22 and the covering markers comprise essentially less than all of the polymorphisms in the region of  
23 interest,

24  
25 criterion b) is the covering markers are substantially nonevenly distributed across a chromosome or a  
26 chromosomal segment,

27  
28 criterion c) is the covering markers are substantially evenly distributed across a chromosome or a  
29 chromosomal segment, and there is a subgroup of one or more of the covering markers, and each of  
30 the markers in the subgroup is chosen without substantial preference for the least common allele  
31 frequency of each of the markers in the subgroup being close to 0.5, and the number of covering  
32 markers in the subgroup is about 5 percent or more of the total number of covering markers,

33  
34 criterion d) is the covering markers are substantially evenly distributed across a chromosome or a  
35 chromosomal segment, and there is a subgroup of one or more of the covering markers, and each of

AMENDED SHEET

72  
98

1 the markers in the subgroup is chosen without substantial preference for the least common allele  
2 frequency of each of the markers in the subgroup being close to 0.5,

3  
4 criterion e) is the covering markers are substantially evenly distributed across a chromosome or a  
5 chromosomal segment, wherein (1) the average chromosomal intermarker distance of the covering  
6 markers is greater than 2 cM or the equivalent thereof and the least common allele frequency of one or  
7 more of the covering markers is less than 0.3, or wherein (2) the least common allele frequency of one  
8 or more of the covering markers is less than 0.2, or wherein (3) the average chromosomal intermarker  
9 distance of the covering markers is greater than 10 cM or the equivalent thereof,

10  
11 criterion f) is the covering markers are substantially evenly distributed across a chromosome or a  
12 chromosomal segment, wherein the average chromosomal intermarker distance of the covering markers  
13 is less than or equal to 2 cM or the equivalent thereof, and wherein the conditional probability the  
14 covering markers were chosen essentially randomly from substantially the known set of bi-allelic  
15 markers with least common allele frequencies between 0.2 inclusive and 0.5 inclusive is less than or  
16 equal to 10 percent, wherein the conditional probability is substantially conditional on (1) the  
17 approximate chromosomal distribution of the covering markers, (2) the marker type of each covering  
18 marker and (3) the CL-F region being N covered to within the CL-F distance  $\delta$  by the two or more bi-  
19 allelic covering markers;

20  
21 criterion g) is the covering markers are substantially evenly distributed across a chromosome or a  
22 chromosomal segment, wherein the average chromosomal intermarker distance of the covering markers  
23 is greater than 2 cM or the equivalent thereof; and wherein the conditional probability the covering  
24 markers were chosen essentially randomly from substantially the known set of bi-allelic markers with  
25 least common allele frequencies between 0.3 inclusive and 0.5 inclusive is less than or equal to 10  
26 percent, wherein the conditional probability is substantially conditional on (1) the approximate  
27 chromosomal distribution of the covering markers, (2) the marker type of each covering marker and (3)  
28 the CL-F region being N covered to within the CL-F distance  $\delta$  by the two or more bi-allelic covering  
29 markers,

30  
31 criterion h) is the covering markers are substantially evenly distributed across a chromosome or a  
32 chromosomal segment, wherein the average chromosomal intermarker distance of the covering markers  
33 is less than or equal to 2 cM or the equivalent thereof; and wherein collection C is essentially the  
34 collection of known groups of bi-allelic markers with least common allele frequencies between 0.2  
35 inclusive and 0.5 inclusive, each group in the collection being substantially similar to the covering

PCT/US99/04376

IPEA/US 22 MAY 2000

73

99

1 markers as a group, wherein a group of bi-allelic markers is a member of collection C if and only if the  
2 group substantially meets criteria (1), (2), (3) and (4); wherein criterion (1) is, each marker in the group  
3 is chosen from substantially the known set of bi-allelic markers with least common allele frequencies  
4 between 0.2 inclusive and 0.5 inclusive, wherein criterion (2) is, the number of markers in the group is  
5 the same as the number of covering markers, wherein criterion (3) is, the chromosomal distribution of  
6 the group of markers and the covering markers is substantially similar, and wherein criterion (4) is, the  
7 marker type of each group marker and the covering marker with substantially the same chromosomal  
8 location is the same; wherein a group that is a member of collection C meets criterion (5) if and only if  
9 (5) the CL-F region is N covered to within a CL-F distance  $\delta$  by the two or more bi-allelic covering  
10 markers; wherein P is essentially the proportion of groups in collection C that meet criterion (5); wherein  
11 P is less than 90 percent;

12  
13 and criterion i) is the covering markers are substantially evenly distributed across a chromosome or a  
14 chromosomal segment, wherein the average chromosomal intermarker distance of the covering markers  
15 is greater than 2 cM or the equivalent thereof, and wherein collection C is essentially the collection of  
16 known groups of bi-allelic markers with least common allele frequencies between 0.3 inclusive and 0.5  
17 inclusive, each group in the collection being substantially similar to the covering markers as a group;  
18 wherein a group of bi-allelic markers is a member of collection C if and only if the group substantially  
19 meets criteria (1), (2), (3) and (4); wherein criterion (1) is, each marker in the group is chosen from  
20 substantially the known set of bi-allelic markers with least common allele frequencies between 0.3  
21 inclusive and 0.5 inclusive, wherein criterion (2) is, the number of markers in the group is the same as  
22 the number of covering markers, wherein criterion (3) is, the chromosomal distribution of the group of  
23 markers and the covering markers is substantially similar, and wherein criterion (4) is, the marker type of  
24 each group marker and the covering marker with substantially the same chromosomal location is the  
25 same, wherein a group that is a member of collection C meets criterion (5) if and only if (5) the CL-F  
26 region is N covered to within a CL-F distance  $\delta$  by the two or more bi-allelic covering markers, wherein P  
27 is essentially the proportion of groups in collection C that meet criterion (5), wherein P is less than 90  
28 percent.

29 75. A use as in claim 74, wherein  $\delta$  is less than or equal to about [1 cM, 0.15] or the equivalent thereof

30 76. A use as in any one of claims 59-75, wherein there is a subgroup of the covering markers, and the  
31 markers in the subgroup are a majority of the covering markers, and each marker in the subgroup is an  
32 SNP, or a bi-allelic marker equivalent formed only from one or more SNPs

33 77. A use as in claim 76 wherein  $L_{MC}$  is less than or equal to about 250,000 bp or the equivalent thereof,  
34  $W_{MC}$  is less than or equal to about 0.15, wherein the species is human being.

**AMENDED SHEET**

74  
~~100~~

1 78. One or more copies of a set of oligonucleotides, the set of oligonucleotides being substantially  
2 complementary to a group of two or more bi-allelic covering markers of the same species, wherein the  
3 group of covering markers systematically cover a CL-F region, the CL-F region being a collection of  
4 points on a two-dimensional plane, the two-dimensional plane having the two orthogonal dimensions of  
5 chromosomal location and least common allele frequency.

6 79. One or more copies of a set of oligonucleotides as in claim 78, wherein the CL-F region is N covered  
7 to within a CL-F distance  $\delta$  by the two or more bi-allelic covering markers, so that each point in the  
8 region is within the CL-F distance  $\delta$  of N or more of the covering markers, wherein  $\delta$  is equal to about [  
9  $\delta_{CL}$ , 0.25] or the equivalent thereof,  $\delta_{CL}$  is equal to the largest chromosomal length, computed by any  
10 method, for which linkage disequilibrium has been observed between any polymorphisms in any  
11 population of the species, N is an integer greater than or equal to 1.

12 80. One or more copies of a set of oligonucleotides as in claim 78, wherein the CL-F region includes 81  
13 percent or more of the centerpoints of the matrix centerpoint lattice of a CL-F matrix, the number of cells  
14 in the matrix being greater than or equal to three, wherein the matrix has R rows and C columns, each  
15 cell of the matrix being of length  $L_{MC}$  and width  $W_{MC}$ , and  $L_{MC}$  being less than or equal to about  $2\delta_{CL}$ , and  
16  $W_{MC}$  being less than or equal to 0.5,  $\delta_{CL}$  is equal to the largest chromosomal length, computed by any  
17 method, for which linkage disequilibrium has been observed between any polymorphisms in any  
18 population of the species, there being N or more covering markers in each cell of 81 percent or more of  
19 the cells of the matrix, N is an integer greater than or equal to 1; the covering markers being distributed  
20 over a chromosomal region of interest, the region of interest being approximately the smallest  
21 chromosome interval that contains all of the covering markers, and the covering markers comprising  
22 essentially less than all of the polymorphisms in the region of interest.

23 81. One or more copies of a set of oligonucleotides as in claim 80, wherein the covering markers are  
24 substantially nonevenly distributed across a chromosome or a chromosomal segment

25 82. One or more copies of a set of oligonucleotides as in claim 80, wherein the covering markers are  
26 substantially evenly distributed across a chromosome or a chromosomal segment, and wherein there is  
27 a subgroup of one or more of the covering markers, and each of the markers in the subgroup is chosen  
28 without substantial preference for the least common allele frequency of each of the markers in the  
29 subgroup being close to 0.5, and the number of covering markers in the subgroup is about 5 percent or  
30 more of the total number of covering markers

31 83. One or more copies of a set of oligonucleotides as in claim 80, wherein the covering markers are  
32 substantially evenly distributed across a chromosome or a chromosomal segment, and wherein there is  
33 a subgroup of one or more of the covering markers, and each of the markers in the subgroup is chosen  
34 without substantial preference for the least common allele frequency of each of the markers in the  
35 subgroup being close to 0.5

**AMENDED SHEET**

75

101

84. One or more copies of a set of oligonucleotides as in claim 80, wherein the covering markers are substantially evenly distributed across a chromosome or a chromosomal segment, wherein (1) the average chromosomal intermarker distance of the covering markers is greater than 2 cM or the equivalent thereof and the least common allele frequency of one or more of the covering markers is less than 0.3, or wherein (2) the least common allele frequency of one or more of the covering markers is less than 0.2, or wherein (3) the average chromosomal intermarker distance of the covering markers is greater than 10 cM or the equivalent thereof

85. One or more copies of a set of oligonucleotides as in claim 80, wherein the covering markers are substantially evenly distributed across a chromosome or a chromosomal segment, wherein the average chromosomal intermarker distance of the covering markers is less than or equal to 2 cM or the equivalent thereof, and wherein the conditional probability the covering markers were chosen essentially randomly from substantially the known set of bi-allelic markers with least common allele frequencies between 0.2 inclusive and 0.5 inclusive is less than or equal to 10 percent; wherein the conditional probability is substantially conditional on (1) the approximate chromosomal distribution of the covering markers, (2) the marker type of each covering marker and (3) there being N or more covering markers in each cell of 81 percent or more of the cells of the matrix

86. One or more copies of a set of oligonucleotides as in claim 80, wherein the covering markers are substantially evenly distributed across a chromosome or a chromosomal segment, wherein the average chromosomal intermarker distance of the covering markers is greater than 2 cM or the equivalent thereof; and wherein the conditional probability the covering markers were chosen essentially randomly from substantially the known set of bi-allelic markers with least common allele frequencies between 0.3 inclusive and 0.5 inclusive is less than or equal to 10 percent; wherein the conditional probability is substantially conditional on (1) the approximate chromosomal distribution of the covering markers, (2) the marker type of each covering marker and (3) there being N or more covering markers in each cell of 81 percent or more of the cells of the matrix.

87. One or more copies of a set of oligonucleotides as in claim 85, wherein the chromosome or the chromosomal segment consists essentially of a set of nonoverlapping chromosome segments of substantially equal length, and wherein each chromosome segment of the set has two or less covering markers located thereon, wherein one and only one covering marker is located on each of 80 percent or more of the chromosome segments of the set, and wherein zero or two and only two covering markers are located on each of 20 percent or less of the chromosome segments of the set, and wherein each chromosome segment that borders a chromosome segment with zero covering markers located thereon has only one or two covering markers located thereon, and wherein each chromosome segment that borders a chromosome segment with two covering markers located thereon has only one or zero covering markers located thereon; and wherein the conditional probability the covering markers were

76  
102

chosen essentially randomly from substantially the known set of bi-allelic markers with least common allele frequencies between 0.2 inclusive and 0.5 inclusive is less than or equal to 10 percent, wherein the conditional probability is substantially conditional on (1) the chromosomal distribution of the covering markers on the chromosome segments of the set, (2) the marker type of each covering marker and (3) there being N or more covering markers in each cell of 81 percent or more of the cells of the matrix 88. One or more copies of a set of oligonucleotides as in claim 86, wherein the chromosome or the chromosomal segment consists essentially of a set of nonoverlapping chromosome segments of substantially equal length, and wherein each chromosome segment of the set has two or less covering markers located thereon, wherein one and only one covering marker is located on each of 80 percent or more of the chromosome segments of the set, and wherein zero or two and only two covering markers are located on each of 20 percent or less of the chromosome segments of the set, and wherein each chromosome segment that borders a chromosome segment with zero covering markers located thereon has only one or two covering markers located thereon, and wherein each chromosome segment that borders a chromosome segment with two covering markers located thereon has only one or zero covering markers located thereon; and wherein the conditional probability the covering markers were chosen essentially randomly from substantially the known set of bi-allelic markers with least common allele frequencies between 0.3 inclusive and 0.5 inclusive is less than or equal to 10 percent, wherein the conditional probability is substantially conditional on (1) the chromosomal distribution of the covering markers on the chromosome segments of the set, (2) the marker type of each covering marker and (3) there being N or more covering markers in each cell of 81 percent or more of the cells of the matrix.

89. One or more copies of a set of oligonucleotides as in claim 80, wherein the covering markers are substantially evenly distributed across a chromosome or a chromosomal segment, wherein the average chromosomal intermarker distance of the covering markers is less than or equal to 2 cM or the equivalent thereof, and wherein collection C is essentially the collection of known groups of bi-allelic markers with least common allele frequencies between 0.2 inclusive and 0.5 inclusive, each group in the collection being substantially similar to the covering markers as a group; wherein a group of bi-allelic markers is a member of collection C if and only if the group substantially meets criteria (1), (2), (3) and (4); wherein criterion (1) is, each marker in the group is chosen from substantially the known set of bi-allelic markers with least common allele frequencies between 0.2 inclusive and 0.5 inclusive, wherein criterion (2) is, the number of markers in the group is the same as the number of covering markers, wherein criterion (3) is, the chromosomal distribution of the group of markers and the covering markers is substantially similar, and wherein criterion (4) is, the marker type of each group marker and the covering marker with substantially the same chromosomal location is the same, wherein a group that is a member of collection C meets criterion (5) if and only if (5) there are N or more of the group markers in each cell of 81 percent or more of the cells of a CL-F matrix with cells having length  $L_{Mc}$  and width  $W_{Mc}$

1 in R rows and C columns; wherein P is essentially the proportion of groups in collection C that meet  
2 criterion (5); wherein P is less than 90 percent  
3 90 One or more copies of a set of oligonucleotides as in claim 80, wherein the covering markers are  
4 substantially evenly distributed across a chromosome or a chromosomal segment, wherein the average  
5 chromosomal intermarker distance of the covering markers is greater than 2 cM or the equivalent  
6 thereof, and wherein collection C is essentially the collection of known groups of bi-allelic markers with  
7 least common allele frequencies between 0.3 inclusive and 0.5 inclusive, each group in the collection  
8 being substantially similar to the covering markers as a group; wherein a group of bi-allelic markers is a  
9 member of collection C if and only if the group substantially meets criteria (1), (2), (3) and (4); wherein  
10 criterion (1) is, each marker in the group is chosen from substantially the known set of bi-allelic markers  
11 with least common allele frequencies between 0.3 inclusive and 0.5 inclusive, wherein criterion (2) is,  
12 the number of markers in the group is the same as the number of covering markers, wherein criterion  
13 (3) is, the chromosomal distribution of the group of markers and the covering markers is substantially  
14 similar, and wherein criterion (4) is, the marker type of each group marker and the covering marker with  
15 substantially the same chromosomal location is the same; wherein a group that is a member of  
16 collection C meets criterion (5) if and only if (5) there are N or more of the group markers in each cell of  
17 81 percent or more of the cells of a CL-F matrix with cells having length  $L_{MC}$  and width  $W_{MC}$  in R rows  
18 and C columns, wherein P is essentially the proportion of groups in collection C that meet criterion (5);  
19 wherein P is less than 90 percent.

20 91 One or more copies of a set of oligonucleotides as in claim 80, wherein the covering markers are  
21 substantially evenly distributed across a chromosome or a chromosomal segment, wherein the average  
22 chromosomal intermarker distance of the covering markers is less than or equal to 2 cM or the  
23 equivalent thereof, wherein the chromosome or the chromosomal segment consists essentially of a set  
24 of nonoverlapping chromosome segments of substantially equal length, and wherein each chromosome  
25 segment of the set has two or less covering markers located thereon, wherein one and only one  
26 covering marker is located on each of 80 percent or more of the chromosome segments of the set, and  
27 wherein zero or two and only two covering markers are located on each of 20 percent or less of the  
28 chromosome segments of the set, and wherein each chromosome segment that borders a chromosome  
29 segment with zero covering markers located thereon has only one or two covering markers located  
30 thereon, and wherein each chromosome segment that borders a chromosome segment with two  
31 covering markers located thereon has only one or zero covering markers located thereon, wherein  
32 collection D is essentially the collection of known groups of bi-allelic markers with least common allele  
33 frequencies between 0.2 inclusive and 0.5 inclusive, each group in the collection being substantially  
34 similar to the covering markers as a group, wherein a group of bi-allelic markers is a member of  
35 collection D if and only if the group substantially meets criteria (1), (2), and (3) (1) each marker in the

IPEA/US 22 MAY 2000

78

104

1 group is chosen from substantially the known set of bi-allelic markers with least common allele  
2 frequencies between 0.2 inclusive and 0.5 inclusive, (2) the number of covering markers and the  
3 number of group markers located on each chromosome segment of the set is the same, and (3) there is  
4 a group marker of the same type as each covering marker located on the same chromosome segment  
5 of the set as each covering marker, wherein a group that is a member of collection D substantially  
6 meets criterion (5) if and only if (5) there are N or more of the group markers in each cell of the matrix,  
7 wherein P is essentially the proportion of groups in collection D that meet criterion (5), wherein P is less  
8 than 90 percent

9 92. One or more copies of a set of oligonucleotides as in claim 80, wherein the covering markers are  
10 substantially evenly distributed across a chromosome or a chromosomal segment, wherein the average  
11 chromosomal intermarker distance of the covering markers is greater than 2 cM or the equivalent  
12 thereof; wherein the chromosome or the chromosomal segment consists essentially of a set of  
13 nonoverlapping chromosome segments of substantially equal length, and wherein each chromosome  
14 segment of the set has two or less covering markers located thereon, wherein one and only one  
15 covering marker is located on each of 80 percent or more of the chromosome segments of the set, and  
16 wherein zero or two and only two covering markers are located on each of 20 percent or less of the  
17 chromosome segments of the set, and wherein each chromosome segment that borders a chromosome  
18 segment with zero covering markers located thereon has only one or two covering markers located  
19 thereon, and wherein each chromosome segment that borders a chromosome segment with two  
20 covering markers located thereon has only one or zero covering markers located thereon; wherein  
21 collection D is essentially the collection of known groups of bi-allelic markers with least common allele  
22 frequencies between 0.3 inclusive and 0.5 inclusive, each group in the collection being substantially  
23 similar to the covering markers as a group; wherein a group of bi-allelic markers is a member of  
24 collection D if and only if the group substantially meets criteria (1), (2), and (3). (1) each marker in the  
25 group is chosen from substantially the known set of bi-allelic markers with least common allele  
26 frequencies between 0.3 inclusive and 0.5 inclusive, (2) the number of covering markers and the  
27 number of group markers located on each chromosome segment of the set is the same, and (3) there is  
28 a group marker of the same type as each covering marker located on the same chromosome segment  
29 of the set as each covering marker; wherein a group that is a member of collection D substantially  
30 meets criterion (5) if and only if (5) there are N or more of the group markers in each cell of the matrix,  
31 wherein P is essentially the proportion of groups in collection D that meet criterion (5), wherein P is less  
32 than 90 percent

33 93. One or more copies of a set of oligonucleotides as in claim 79, wherein the covering markers meet  
34 criterion a) and one or more of the criteria b), c), d), e), f), g), h) or i), wherein  
35

**AMENDED SHEET**



79

~~105~~

1 criterion a) is the covering markers are distributed over a chromosomal region of interest, the region of  
2 interest being approximately the smallest chromosome interval that contains all of the covering markers,  
3 and the covering markers comprise essentially less than all of the polymorphisms in the region of  
4 interest;

5  
6 criterion b) is the covering markers are substantially nonevenly distributed across a chromosome or a  
7 chromosomal segment;

8  
9 criterion c) is the covering markers are substantially evenly distributed across a chromosome or a  
10 chromosomal segment, and there is a subgroup of one or more of the covering markers, and each of  
11 the markers in the subgroup is chosen without substantial preference for the least common allele  
12 frequency of each of the markers in the subgroup being close to 0.5, and the number of covering  
13 markers in the subgroup is about 5 percent or more of the total number of covering markers,

14  
15 criterion d) is the covering markers are substantially evenly distributed across a chromosome or a  
16 chromosomal segment, and there is a subgroup of one or more of the covering markers, and each of  
17 the markers in the subgroup is chosen without substantial preference for the least common allele  
18 frequency of each of the markers in the subgroup being close to 0.5,

19  
20 criterion e) is the covering markers are substantially evenly distributed across a chromosome or a  
21 chromosomal segment, wherein (1) the average chromosomal intermarker distance of the covering  
22 markers is greater than 2 cM or the equivalent thereof and the least common allele frequency of one or  
23 more of the covering markers is less than 0.3, or wherein (2) the least common allele frequency of one  
24 or more of the covering markers is less than 0.2, or wherein (3) the average chromosomal intermarker  
25 distance of the covering markers is greater than 10 cM or the equivalent thereof;

26  
27 criterion f) is the covering markers are substantially evenly distributed across a chromosome or a  
28 chromosomal segment, wherein the average chromosomal intermarker distance of the covering markers  
29 is less than or equal to 2 cM or the equivalent thereof, and wherein the conditional probability the  
30 covering markers were chosen essentially randomly from substantially the known set of bi-allelic  
31 markers with least common allele frequencies between 0.2 inclusive and 0.5 inclusive is less than or  
32 equal to 10 percent, wherein the conditional probability is substantially conditional on (1) the  
33 approximate chromosomal distribution of the covering markers, (2) the marker type of each covering  
34 marker and (3) the CL-F region being N covered to within the CL-F distance  $\delta$  by the two or more bi-  
35 allelic covering markers;

**AMENDED SHEET**

1  
2 criterion g) is the covering markers are substantially evenly distributed across a chromosome or a  
3 chromosomal segment, wherein the average chromosomal intermarker distance of the covering markers  
4 is greater than 2 cM or the equivalent thereof, and wherein the conditional probability the covering  
5 markers were chosen essentially randomly from substantially the known set of bi-allelic markers with  
6 least common allele frequencies between 0.3 inclusive and 0.5 inclusive is less than or equal to 10  
7 percent, wherein the conditional probability is substantially conditional on (1) the approximate  
8 chromosomal distribution of the covering markers, (2) the marker type of each covering marker and (3)  
9 the CL-F region being N covered to within the CL-F distance  $\delta$  by the two or more bi-allelic covering  
10 markers;

11  
12 criterion h) is the covering markers are substantially evenly distributed across a chromosome or a  
13 chromosomal segment, wherein the average chromosomal intermarker distance of the covering markers  
14 is less than or equal to 2 cM or the equivalent thereof; and wherein collection C is essentially the  
15 collection of known groups of bi-allelic markers with least common allele frequencies between 0.2  
16 inclusive and 0.5 inclusive, each group in the collection being substantially similar to the covering  
17 markers as a group; wherein a group of bi-allelic markers is a member of collection C if and only if the  
18 group substantially meets criteria (1), (2), (3) and (4); wherein criterion (1) is, each marker in the group  
19 is chosen from substantially the known set of bi-allelic markers with least common allele frequencies  
20 between 0.2 inclusive and 0.5 inclusive, wherein criterion (2) is, the number of markers in the group is  
21 the same as the number of covering markers, wherein criterion (3) is, the chromosomal distribution of  
22 the group of markers and the covering markers is substantially similar, and wherein criterion (4) is, the  
23 marker type of each group marker and the covering marker with substantially the same chromosomal  
24 location is the same, wherein a group that is a member of collection C meets criterion (5) if and only if  
25 (5) the CL-F region is N covered to within a CL-F distance  $\delta$  by the two or more bi-allelic covering  
26 markers, wherein P is essentially the proportion of groups in collection C that meet criterion (5); wherein  
27 P is less than 90 percent,

28  
29 and criterion i) is the covering markers are substantially evenly distributed across a chromosome or a  
30 chromosomal segment, wherein the average chromosomal intermarker distance of the covering markers  
31 is greater than 2 cM or the equivalent thereof, and wherein collection C is essentially the collection of  
32 known groups of bi-allelic markers with least common allele frequencies between 0.3 inclusive and 0.5  
33 inclusive, each group in the collection being substantially similar to the covering markers as a group,  
34 wherein a group of bi-allelic markers is a member of collection C if and only if the group substantially  
35 meets criteria (1), (2), (3) and (4); wherein criterion (1) is, each marker in the group is chosen from

81  
107

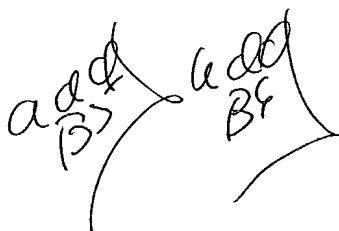
1 substantially the known set of bi-allelic markers with least common allele frequencies between 0.3  
2 inclusive and 0.5 inclusive, wherein criterion (2) is, the number of markers in the group is the same as  
3 the number of covering markers, wherein criterion (3) is, the chromosomal distribution of the group of  
4 markers and the covering markers is substantially similar, and wherein criterion (4) is, the marker type of  
5 each group marker and the covering marker with substantially the same chromosomal location is the  
6 same, wherein a group that is a member of collection C meets criterion (5) if and only if (5) the CL-F  
7 region is N covered to within a CL-F distance  $\delta$  by the two or more bi-allelic covering markers, wherein P  
8 is essentially the proportion of groups in collection C that meet criterion (5), wherein P is less than 90  
9 percent

10 94. One or more copies of a set of oligonucleotides as in claim 93, wherein  $\delta$  is less than or equal to  
11 about [1 cM, 0.15] or the equivalent thereof.

12 95. One or more copies of a set of oligonucleotides as in any one of claims 78-94, wherein there is a  
13 subgroup of the covering markers, and the markers in the subgroup are a majority of the covering  
14 markers, and each marker in the subgroup is an SNP, or a bi-allelic marker equivalent formed only from  
15 one or more SNPs

16 96. One or more copies of a set of oligonucleotides as in claim 95 wherein  $L_{MC}$  is less than or equal to  
17 about 250,000 bp or the equivalent thereof,  $W_{MC}$  is less than or equal to about 0.15, wherein the species  
18 is human being

19  
20  
21  
22

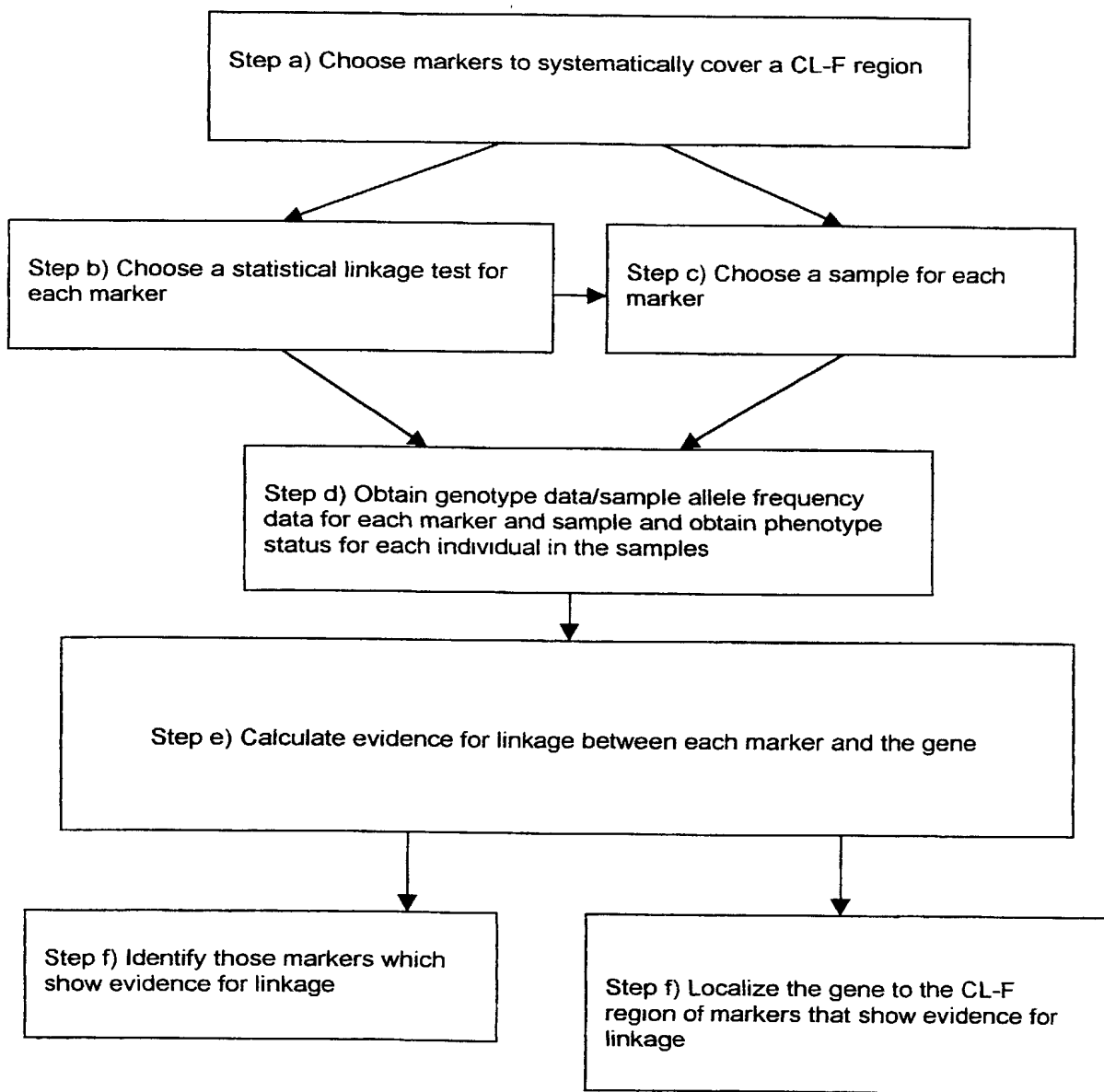
**AMENDED SHEET**

**Abstract**

Versions of the invention are directed to methods (including software), apparatus, compositions of matter, and new uses of compositions of matter for a new type of association based linkage study technique using bi-allelic markers. In this new type of association based linkage study technique, the bi-allelic markers used in the new linkage studies are chosen so that the least common allele frequencies of the markers vary systematically over a range or subrange of least common allele frequency and the chromosomal location of the markers vary systematically over one or more chromosomes or chromosomal regions. And the bi-allelic markers are chosen so that the markers' chromosomal locations and least common allele frequencies vary systematically in an essentially independent manner. This selection of markers achieves a systematic distribution of the markers over a two-dimensional region having the orthogonal dimensions of chromosomal location and least common allele frequency. By using the two characteristics or two dimensions of marker chromosomal location and marker allele population frequency in this unique way, the power and systematic nature of genetic linkage studies using association based linkage tests is greatly increased. These unique two-dimensional linkage study techniques increase the power of association based linkage studies to localize trait causing genes or polymorphisms of modest effect such as human disease causing polymorphisms.

1/1

**Drawing #1: Computer Program Flowsheet for Process #1 and Process#1A**



Please type a plus sign (+) in this box ☐PTO/SB/01 (12-97)  
Approved for use through 9/30/00 OMB 0651-0032

Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number

**DECLARATION FOR UTILITY OR  
DESIGN  
PATENT APPLICATION  
(37 CFR 1.63)**☒ Declaration  
Submitted with Initial  
Filing OR ☐ Declaration  
Submitted after Initial  
Filing (surcharge  
(37 CFR 1.16 (e))  
required)

Attorney Docket Number

First Named Inventor

MCGINNIS, Ralph

**COMPLETE IF KNOWN**

Application Number

PCT/US99/04376

Filing Date

26 February 1999

Group Art Unit

Examiner Name

Horlick, Kenneth

As a below named inventor, I hereby declare that:

My residence, post office address, and citizenship are as stated below next to my name

I believe I am the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if plural names are listed below) of the subject matter which is claimed and for which a patent is sought on the invention entitled

**TWO-DIMENSIONAL LINKAGE STUDY TECHNIQUES**

the specification of which

(Title of the Invention)

☐ is attached hereto  
OR☒ was filed on (MM/DD/YYYY) **02/26/1999** as United States Application Number or PCT InternationalApplication Number **PCT/US99/04376** and was amended on (MM/DD/YYYY) **04/17/2000** (if applicable).

I hereby state that I have reviewed and understand the contents of the above identified specification, including the claims, as amended by any amendment specifically referred to above

I acknowledge the duty to disclose information which is material to patentability as defined in 37 CFR 1.56

I hereby claim foreign priority benefits under 35 U.S.C. 119(a)-(d) or 365(b) of any foreign application(s) for patent or inventor's certificate, or 365(a) of any PCT international application which designated at least one country other than the United States of America, listed below and have also identified below, by checking the box, any foreign application for patent or inventor's certificate, or of any PCT international application having a filing date before that of the application on which priority is claimed.

Prior Foreign Application Number(s)	Country	Foreign Filing Date (MM/DD/YYYY)	Priority Not Claimed	Certified Copy Attached?	
				YES	NO
			<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
			<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
			<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
			<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

☐ Additional foreign application numbers are listed on a supplemental priority data sheet PTO/SB/02B attached hereto

I hereby claim the benefit under 35 U.S.C. 119(e) of any United States provisional application(s) listed below

Application Number(s)	Filing Date (MM/DD/YYYY)	<input type="checkbox"/> Additional provisional application numbers are listed on a supplemental priority data sheet PTO/SB/02B attached hereto
US 60/076,102	02/26/1998	
US 60/076,182	02/27/1998	
US 60/086,947	05/27/1998	
US 60/107,673	11/07/1998	

[Page 1 of 2]

Burden Hour Statement: This form is estimated to take 0.4 hours to complete. Time will vary depending upon the needs of the individual case. Any comments on the amount of time you are required to complete this form should be sent to the Chief Information Officer, Patent and Trademark Office, Washington, DC 20231. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Assistant Commissioner for Patents, Washington, DC 20231.

Please type a plus sign (+) inside this box → ☐

PTO/SB/01 (12-97)  
 Approved for use through 9/30/00. OMB 0851-0032  
 Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

## DECLARATION — Utility or Design Patent Application

I hereby claim the benefit under 35 U.S.C. 120 of any United States application(s), or 365(c) of any PCT international application designating the United States of America, listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in the prior United States or PCT international application in the manner provided by the first paragraph of 35 U.S.C. 112, I acknowledge the duty to disclose information which is material to patentability as defined in 37 CFR 1.56 which became available between the filing date of the prior application and the national or PCT international filing date of this application.

U.S. Parent Application or PCT Parent Number	Parent Filing Date (MM/DD/YYYY)	Parent Patent Number (if applicable)
PCT/US99/04376	26 Feb 1999	

☐ Additional U.S. or PCT international application numbers are listed on a supplemental priority data sheet PTO/SB/02B attached hereto.

As a named inventor, I hereby appoint the following registered practitioner(s) to prosecute this application and to transact all business in the Patent and Trademark Office connected therewith:

Customer Number		Registered practitioner(s) name/registration number listed below	
Name	Registration Number	Name	Registration Number
Robert McGinnis	44, 232		

☐ Additional registered practitioner(s) named on supplemental Registered Practitioner Information sheet PTO/SB/02C attached hereto.

Direct all correspondence to: ☐ Customer Number or Bar Code Label ☒ Correspondence address below

Name	Robert McGinnis		
Address	1575 West Kagy Blvd.		
Address			
City	Bozeman	State	MT
ZIP	59715		
Country	USA	Telephone	406-522-9355
Fax			

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under 18 U.S.C. 1001 and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

Name of Sole or First Inventor: ☐ A petition has been filed for this unsigned inventor

Given Name (first and middle (if any))		Family Name or Surname	
Ralph Evan		MCGINNIS	
Inventor's Signature	Ralph Evan McGinnis		Date
			8/25/00
Residence: City	Bishop's Startford	State	
		Country	G. Britain
Post Office Address	26 Southmill Road		
Post Office Address	Bishop's Startford Herts CM233DP (England) Great Britain		
City		State	
		ZIP	
		Country	

☒ Additional inventors are being named on the supplemental Additional Inventor(s) sheet(s) PTO/SB/02A attached hereto

174623068 1032600

Please type a plus sign (+) inside this box: ☐

PTO/SB/02A (3-97)

Approved for use through 9/30/98 OMB 0651-0032  
Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE  
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

+

## DECLARATION

## ADDITIONAL INVENTOR(S) Supplemental Sheet

Page 1 of 1

Name of Additional Joint Inventor, if any:

☐ A petition has been filed for this unsigned inventor

Given Name (first and middle [if any])

Family Name or Surname

Robert, Owen

MCGINNIS

Inventor's  
Signature

Robert McGinnis

Date

08/25/2000

Residence: City

Bozeman

State

MT

Country

USA

Citizenship

USA

Post Office Address

1575 West Kagy Blvd.

Post Office Address

City

Bozeman

State

MT

ZIP

59715

Country

USA

Name of Additional Joint Inventor, if any:

☐ A petition has been filed for this unsigned inventor

Given Name (first and middle [if any])

Family Name or Surname

Inventor's  
Signature

Date

Residence: City

State

Country

Citizenship

Post Office Address

Post Office Address

City

State

ZIP

Country

Name of Additional Joint Inventor, if any:

☐ A petition has been filed for this unsigned inventor

Given Name (first and middle [if any])

Family Name or Surname

Inventor's  
Signature

Date

Residence: City

State

Country

Citizenship

Post Office Address

Post Office Address

City

State

ZIP

Country

Burden Hour Statement This form is estimated to take 0.4 hours to complete. Time will vary depending upon the needs of the individual case. Any comments on the amount of time you are required to complete this form should be sent to the Chief Information Officer, Patent and Trademark Office, Washington, DC 20231. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Assistant Commissioner for Patents, Washington, DC 20231

+



**STATEMENT CLAIMING SMALL ENTITY STATUS  
(37 CFR 1.9(f) & 1.27(b))--INDEPENDENT INVENTOR**

Applicant, Patentee, or Identifier: MCGINNIS, Ralph, Evan  
Application or Patent No.: PCT/US99/04376  
Filed or Issued: 26 February 1999  
Title: TWO-DIMENSIONAL LINKAGE STUDY TECHNIQUES

☐ the specification filed herewith with title as listed above.

☒ the application identified above.

☐ the patent identified above.

Each person, concern, or organization to which I have assigned, granted, conveyed, or licensed or am under an obligation under contract or law to assign, grant, convey, or license any rights in the invention is listed below:

- ☒ No such person, concern, or organization exists.
- ☐ Each such person, concern, or organization is listed below.

I acknowledge the duty to file, in this application or patent, notification of any change in status resulting in loss of entitlement to small entity status prior to paying, or at the time of paying, the earliest of the issue fee or any maintenance fee due after the date on which status as a small entity is no longer appropriate. (37 CFR 1.28(b))

<u>Ralph E McGinnis</u> NAME OF INVENTOR	<u>NAME OF INVENTOR</u>	<u>NAME OF INVENTOR</u>
<u>Ralph E McGinnis</u> Signature of inventor	<u>Signature of inventor</u>	<u>Signature of inventor</u>
<u>August 9, 2000</u> Date	<u>Date</u>	<u>Date</u>

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

**STATEMENT CLAIMING SMALL ENTITY STATUS  
(37 CFR 1.9(f) & 1.27(b))--INDEPENDENT INVENTOR**

Docket Number (Optional)

**MCGINNIS, Robert, Owen**

Applicant, Patentee, or Identifier:

**PCT/US99/04376**

Application or Patent No.:

**26 February 1999**

**Filed or Issued:**

## TWO-DIMENSIONAL LINKAGE STUDY TECHNIQUES

As a below named inventor, I hereby state that I qualify as an independent inventor as defined in 37 CFR 1.9(c) for purposes of paying reduced fees to the Patent and Trademark Office described in:

- ☐ the specification filed herewith with title as listed above.
- ☒ the application identified above.
- ☐ the patent identified above.

I have not assigned, granted, conveyed, or licensed, and am under no obligation under contract or law to assign, grant, convey, or license, any rights in the invention to any person who would not qualify as an independent inventor under 37 CFR 1.9(c) if that person had made the invention, or to any concern which would not qualify as a small business concern under 37 CFR 1.9(d) or a nonprofit organization under 37 CFR 1.9(e).

Each person, concern, or organization to which I have assigned, granted, conveyed, or licensed or am under an obligation under contract or law to assign, grant, convey, or license any rights in the invention is listed below:

- ☒ No such person, concern, or organization exists.
- ☐ Each such person, concern, or organization is listed below

Separate statements are required from each named person, concern, or organization having rights to the invention stating their status as small entities. (37 CFR 1.27)

I acknowledge the duty to file, in this application or patent, notification of any change in status resulting in loss of entitlement to small entity status prior to paying, or at the time of paying, the earliest of the issue fee or any maintenance fee due after the date on which status as a small entity is no longer appropriate (37 CFR 1.28(b))

McGinnis, Robert  
NAME OF INVENTOR

NAME OF INVENTOR

NAME OF INVENTOR

NAME OF INVENTOR

Robert McGinnis

Signature of inventor

Signature of inventor

Signature of inventor

8/25/2000  
Date

Date \_\_\_\_\_

Date \_\_\_\_\_

Date \_\_\_\_\_